Marker Development for the Durable Leaf Rust Resistance Gene *Lr34* of Wheat Using Sequence Information from rice, *Aegilops tauschii* and *Brachypodium sylvaticum*

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universtät Zürich

von

Eligio Bossolini

aus

Italien

Promotionskomitee

Prof. Dr. Beat Keller (Vorsitz)

Prof. Dr. Enrico Martinoia

Prof. Dr. Ueli Grossniklaus

Zürich 2006

Die vorliegende Arbeit wurde von der Mathematisch-naturwissenschaftlichen Fakultät der Universität Zürich im Wintersemester 2006/2007 als Dissertation angenommen.

Promtionskomitee: Prof. Dr. Beat Keller (Vorsitz), Prof. Dr. Enrico Martinoia, Prof. Dr. Ueli Grossniklaus.

TABLE OF CONTENTS

1. SUMMARY

This thesis describes the development of molecular markers specific for the partial leaf rust *(Puccinia triticina* Eriks.) resistance gene *Lr34* in hexaploid wheat using comparative genomics with sequence information of the diploid goat grass *(Aegilops tauschii)*, rice *(Oryza sativa)* and the wild grass *Brachypodium sylvaticum.* QTL mapping of leaf rust resistance and the *Lr34* gene was previously performed on a population of two-hundred and forty single seed descent (SSD)-derived ($F_{5:7}$) lines from the intraspecific cross between the two adapted Swiss bread wheat (*Triticum aestivum* L) cultivars 'Arina' and 'Forno'. Based on these studies, a major leaf rust resistance QTL (*QLrP.sfr-7DS*) was identified on chromosome 7DS in the winter wheat cultivar 'Forno'. *QLrP.sfr-7DS* conferred a durable and slow-rusting resistance phenotype, co-segregated with a QTL for leaf tip necrosis (LTN) and was mapped close to *Xgwm295* at a very similar location as the adult plant leaf rust resistance gene *Lr34* found in some spring wheat lines of the CIMMYT germplasm. In order to saturate with more markers this agronomically important locus, the '*Lr34* orthologous region' was identified in rice chromosome 6S in a subtelomeric region of ~300kb. The genes of this region of the rice genome were then used to identify the corresponding wheat expressed sequence tags (ESTs). All these orthologous ESTs were monomorphic in the 7D chromosome of the lines 'Arina' and 'Forno'. To get access to more polymorphic intergenic regions, the same ESTs were used to screen the bacterial artificial chromosome (BAC) library of wheat cultivar 'Glenlea' and the BAC library of the diploid grass *Ae. tauschii.* BAC clones containing the ESTs were identified. Those containing several ESTs were considered orthologous to the rice region and were sequenced. Sequence analysis revealed the presence of ten microsatellite sequences which were converted into simple sequence repeat (SSR) markers for genotyping. Three of them were polymorphic in the cross 'Arina' x 'Forno'. Mapping data revealed a local disruption in collinearity between wheat chromosome 7DS and rice chromosome 6S, caused

by an inversion of 220 kb in rice. One of the newly developed microsatellite markers, SWM10, showed the same allele in the three independent sources of *Lr34:* 'Frontana', 'Chinese Spring' and 'Forno', as well in most of the genotypes containing *Lr34*. Therefore, SWM10 is a highly useful marker to assist selection for *Lr34* in breeding programs worldwide. The inversion between rice and wheat in the *Lr34* genomic region hindered marker saturation of the confidence interval using a wheat-rice comparative genomics strategy. Recently, species of the genus *Brachypodium (B. sylvaticum and B. distachyon)* were suggested to provide a better model than rice to study the genome of temperate cereals. Given its phylogenetic position between rice and the Triticeae crops and its compact genome, *Brachypodium* was proposed to bridge the gap between rice and the temperate cereals (e.g. wheat, barley, rye). Thus, the disruption in collinearity between rice and wheat genomes in the *Lr34* region provided a good case study to investigate rice-*Brachypodium*-wheat relationships. We sequenced a 371 kb region in *B. sylvaticum*, and we compared it with orthologous regions from rice and wheat. *Brachypodium* and wheat showed perfect macro-colinearity of genetic markers, whereas rice contained the ~220 kb inversion. Through comparative annotation we identified alternative transcripts and improved the annotation for several rice genes, indicating that ~15% of rice genes might require re-annotation. Surprisingly, our data suggest that 10-15% of functional sequences in small grass genomes may not encode any proteins. From available genomic and EST sequences we estimated *Brachypodium* to have diverged from wheat about 35-40 MYA, significantly less than the divergence of rice and wheat (50 MYA). Our data confirmed that the *B. sylvaticum* genome is more similar to the wheat genome than to the rice genome both in terms of gene order and gene sequence. Nevertheless, the differences in gene content found between *Brachypodium* and wheat indicate that the *Brachypodium* genome will not be a perfect model for the large Triticeae genomes.

2. ZUSAMMENFASSUNG

Diese Arbeit beschreibt die Entwicklung von spezifischen molekularen Markern für das partielle Braunrostresistenzgen *Lr34* in hexaploidem Brotweizen unter Verwendung vergleichender Genomik und Sequenzinformationen aus dem diploiden Ziegengras (*Aegilops tauschii*), Reis (*Oryza sativa*) und der Wald-Fiederzwenke (*Brachypodium sylvaticum*). Die QTL Kartierung der Braunrostresistenz und von *Lr34* wurde bereits in einer früheren Arbeit durchgeführt, an einer Population von zweihundertundvierzig rekombinanten Inzuchtlinien aus der interspezifischen Kreuzung zweier adaptierter Schweizer Brotweizensorten 'Arina' und 'Forno', über Single Seed Descent (SSD) bis zur $F_{5:7}$ Generation erzeugt. Basierend auf diesen Studien wurde in der Winterweizensorte 'Forno' ein bedeutender Braunrost-QTL (*QLrP.sfr-7DS*) auf dem Chromosom 7DS identifiziert. *QLrP.sfr-7DS* verlieh einen dauerhaften „slow –rusting" Resistenz-Phänotyp, kosegregierte mit einem QTL für Blattspitzendürre (LTN) und wurde an einer ähnlichen Stelle wie das Braunrostresistenzgen *Lr34*, welches in einigen Sommerweizensorten des CIMMYT gefunden wurde, nahe des Markers *Xgwm295* kartiert. Um diesen agronomisch wichtigen Lokus noch weiter mit molekularen Markern saturieren zu können, wurde die orthologe Region zu *Lr34* in Reis in einer 300 kb grossen subtelomeren Region auf Chromosom 6 identifiziert. Die Gene dieser Genomregion aus Reis wurden dann verwendet, um die entsprechenden Weizen ESTs zu ermitteln. All diese orthologen ESTs waren monomorph im 7D Chromosom der Sorten 'Arina' und 'Forno'. Um Zugang zu den polymorpheren, intergenischen Regionen zu bekommen, wurden die selben ESTs dazu verwendet, um Bakterielle artifizielle Chromosomen (BAC)- Bibliotheken der Weizensorte 'Glenlea' und des diploiden Grases *Ae.tauschii* zu durchsuchen. BAC-Klone, welche die ESTs enthielten, wurden identifiziert. Jene, welche mehrere ESTs enthielten, wurden als ortholog zur Reisregion betrachtet und sequenziert. Die Sequenzanalyse enthüllte zehn Mikrosatelliten-Sequenzen, welche für die

Genotypisierung in „simple sequence repeat“ (SSR) Marker umgewandelt wurden. Drei dieser Marker waren polymorph in der Kreuzung ‘Arina’ x ‘Forno’. Die Kartierungsdaten deckten einen lokalen Unterbruch in der Kolinearität zwischen dem Weizenchromosom 7DS und Reischromosom 6 auf, welche durch eine 220 kb grosse Inversion in Reis verursacht wurde. Einer der kürzlich entwickelten Mikrosatelliten-Marker, SWM10, zeigte das gleiche Allel in den drei unabhängigien Quellen von *Lr34*: ‘Frontana’, ‘Chinese Spring’ und ‘Forno’, sowie in den meisten Genotypen, welche *Lr34* enthalten. Deshalb ist SWM10 ein sehr nützlicher Marker, um in Züchtungsprogrammen weltweit bei der Selektion für *Lr34* zu helfen. Die Inversion zwischen Reis und Weizen in der genomischen Region von *Lr34* verhinderte die Saturierung des Konfidenzintervalls mit Markern unter Verwendung einer „Weizen-Reis komparativen genomischen Strategie“. Kürzlich wurde vorgeschlagen, dass Arten der Gattung *Brachypodium* (*B. sylvaticum* und *B. distachyon*) ein besseres Modell für die Genomforschung an Getreide darstellen als Reis. Unter Berücksichtigung seiner phylogenetischen Position zwischen Reis und den Triticeae und seines kompakten Genoms soll *Brachypodium* die Lücke zwischen Reis und den Getreiden aus den gemässigten Klimazonen (z.B. Weizen, Gerste und Roggen) überbrücken. Deshalb bot der Unterbruch in der Kolinearität zwischen den Reis- und Weizengenomen in der *Lr34* Region eine gute Gelegenheit, die Reis-*Brachypodium*-Weizen Beziehung zu untersuchen. Wir sequenzierten eine 371 kb grosse Region in *B. sylvaticum*, welche mit den orthologen Regionen von Reis und Weizen verglichen wurde. *Brachypodium* und Weizen zeigten eine perfekte Makrokolinearität der genetischen Marker, wohingegen Reis die 220 kb grosse Inversion enthielt. Durch vergleichende Annotation identifizierten wir alternative Transkripte und verbesserten die Annotation für mehrere Reisgene. Wir schätzen dass ~15% der Reisgene neu annotiert werden müssten. Überraschenderweise ergaben unsere Daten, dass 10-15% der funktionellen Sequenzen in kleinen Grassgenomen möglicherweise keine Proteine kodieren. Anhand vorhandener genomischer Sequenzen und ESTs schätzen wir, dass *Brachypodium*

und Weizen vor etwa 35-40 Millionen Jahren divergiert sind, massgeblich früher als die Trennung von Reis und Weizen (50 Millionen Jahre). Unsere Daten bestätigten, dass das Genom von *B. sylvaticum* dem Weizengenom ähnlicher ist als dem Reisgenom, sowohl in der Genreihenfolge als auch den Gensequenzen. Dennoch deuten die Unterschiede im Gengehalt zwischen *Brachypodium* und Weizen an, dass das *Brachypodium* Genom nicht das perfekte Modele für die grossen Triticeae Genome sein wird.

Chapter 3

GENERAL INTRODUCTION

3. GENERAL INTRODUCTION

3.1 History of wheat cultivation and present challenges

History of wheat cultivation

Wheat was one of the first domesticated crops and its cultivation started in the Fertile Crescent ~12,000 years ago. The first cultivated wheats where diploid einkorn *(Triticum monococcum* L. spp. *monococcum*, AA genome, 2n = 14) and tetraploid emmer *(T. turgidum* L. spp. *dicoccum* Schübl., AABB genome, 2n = 4x = 28). Nowadays, wheat differs very much from the wild accessions that were initially domesticated. Most of the observed phenotypic variation is probably due to single nucleotide polymorphisms (SNPs) in the coding portion of the genome; thus, a very small portion of the genome is responsible for these differences. The favourable alleles that increase yield, improve stress resistance and specify crop architecture have become more and more abundant in the cultivated germplasm after domestication. Together with the improvement of agronomic techniques, the cultivated local wheat landraces went through a constant and dramatic genetic improvement. Yield of land races increased with a breeding and selection process which is as old as agriculture. However, it is only after 1950 that the yield increased to the values of modern agriculture. The yield increase was achieved with consistent mechanic and chemical input and the release of better performing and shorter varieties. This improvement was so important that nowadays we refer to it as the 'green revolution'. Unfortunately with the advent of the green revolution many of the low-

yielding land races were not cultivated anymore. Many of these land races were stored in seed banks, but the diversity of the cultivated germplasm was generally reduced.

Today, due to population increase and land degradation, a further improvement in agriculture is essential. This improvement will rely on a more efficient exploitation of the available genetic resources, and will be enabled by recently developed techniques of molecular biology and recombinant DNA. This foreseeable change is now called 'the gene revolution'.

Importance of wheat and present challenges for yield increase

Wheat is the most important staple food in the western part of the world. Hexaploid wheat *(Triticum aestivum)* is mostly cultivated for making baked-products, while tetraploid wheat *(Triticum durum)* is commonly used for pasta. In marginal rural areas also the low yield diploid einkorn *(Triticum monococcum)* and tetraploid emmer *(Triticum turgidum)* have still some importance. Wheat is the 1st most widely cultivated crop worldwide and the 2nd most important crop in terms of yield after maize. Considering the cultivated area in the year 2005, it was grown on 217 million hectares (mha) (http://faostat.fao.org/site/408/default.aspx), while rice cultivation covered 154 mha and maize 148 mha. In the same year, 629 million tons (mt) of wheat grain were harvested, 701 mt of maize and 618 mt of rice. Food supply for a constantly growing world population is a challenge for modern agriculture. An environmentally sustainable yield increase has to be achieved without a further expansion of cultivated land. As reported by Oerke and Dehne (2004) "…(the) land suitable for agricultural production is limited, and most of the soils with high productivity potential are already under cultivation". Recurrent drought and salinity problems in marginal agricultural areas prevent a further extension of the world wheat acreage. Given these limitations, wheat yield can mostly

be increased by minimizing crop losses caused by diseases. It is estimated that pathogens and pests together cause a 15-18% yield loss every year worldwide (Oerke and Dehne, 2004). Most of wheat diseases are caused by fungi; some of them are also compromising crop quality by contaminating the grain with toxins. A breeding program focused on the development of wheat lines expressing genetic resistance is an environmentally safe and effective strategy to contain disease epidemics and improve harvest quality.

The modern societies would have developed very differently if agriculture would still require the same human labour input as in the 19th century. It is an important task of our society to keep improving crop yield and quality in a way which is sustainable in a long perspective.

3.2 Genetic resistance to leaf rust in wheat

Wheat rust diseases

Rusts are some of the most frequently occurring diseases of wheat. They are caused by obligate biotrophic fungi of the genus *Puccinia* (class of Basidiomycetes). The spores of the fungus are spread by the wind and germinate on the surface of leaves and stems. With favourable humid - warm conditions they can cause severe epidemics with serious yield losses. Wheat is affected by three distinct rust diseases: leaf (or brown) rust is caused by the fungus *P. triticina* Eriks., stripe (or yellow) rust is caused by the fungus *P. striiformis* Westend. f. sp. *tritici*, and the stem (or black) rust is produced by the fungus *P. graminis* Pers. f. sp. *tritici.* While leaf and stripe rust are endemic in all wheat growing areas, stem rust has only rarely been observed in Europe.

Genetics of leaf rust resistance

Among the rust diseases, leaf rust is the one which occurs most frequently and is most widely spread (Kolmer 1996). Usual losses are between 5-15% of the potential yield (McIntosh et al. 1995). Losses are higher in case of early infection and humid, warm and windy conditions which are beneficial for the spreading of the epidemic. In most parts of the world, control of leaf rust by fungicide treatment is not economically advantageous and deployment of wheat lines having genetic resistance is the most convenient method to limit losses. More than 50 genes for leaf rust resistance (*Lr*) have been described (reviewed by McIntosh et al. 1995), and most of them follow the 'gene for gene' relationship proposed by Flor (1942) which is characteristic of most vertical resistance genes. Unfortunately, vertical resistance genes are easily overcome by mutations or recombinations in the pathogen population. In contrast, horizontal resistance genes are extremely valuable because they do not loose effectiveness; most of them are only partial, but when they are used in combination with vertical resistance genes they confer a level of protection close to immunity. Two *Lr* genes (*Lr34* and *Lr46*) confer durable horizontal resistance in wheat. Both of them co-segregate with the morphological marker leaf tip necrosis (LTN), which is particularly pronounced in the flag leaves and appears after flowering. LTN penetrance is influenced by the environment and its expression depends also on the genetic background of the cultivar. A moderate decrease in yield was observed in near isogenic lines with and without *Lr34*, probably due to LTN; however this yield penalty was more than compensated by the resistance in case of heavy leaf rust infection (Singh and Huerta-Espino 1997). The *Lr34* and *Lr46* genes are defined as slow rusting because they delay the development of the infection (Singh and Huerta-Espino 2003). *Lr34* is located on chromosome 7DS (Dick 1987; Singh et al. 2000; Boukhatem et al. 2002; Suenaga et al. 2003; Ramburan et al. 2004; Schnurbusch et al. 2004a-b) and is one of the most relevant genes in wheat disease resistance breeding. It was originally detected in the spring

wheat material at the CIMMYT (Singh 1992a, Singh and Rajaram 1992). In the Swiss winter wheat cv. 'Forno' we have identified a QTL for leaf rust resistance, *QLrP.sfr-7DS*, which has a very similar mapping location and phenotype as *Lr34* (Schnurbusch et al. 2004a). Therefore, even though no records are available to link the pedigree of 'Forno' to the spring wheat germplasm of the CIMMYT, *QLrP.sfr-7DS* and *Lr34* are possibly the same gene. *Lr34* is a peculiar gene because it confers resistance to multiple diseases, e.g. it was reported to be effective against stripe rust (in this case it is called *Yr18*, Dyck et al. 1966; Singh and Rajaram 1992; Ma and Singh 1996), stem rust (Dyck 1987, Kerber and Aung 1999), powdery mildew (*Blumeria graminis* f. sp. tritici, Spielmeyer et al. 2005) and barley yellow dwarf (BYD) virus (Singh 1993; Ayala et al 2002). A certain degree of resistance was also observed against spot blotch in lines having LTN, even though this was not directly linked with the presence of *Lr34* (Joshi et al. 2004). This effect on multiple diseases makes *Lr34* a unique resource for breeding, and a model to understand the molecular basis of partial horizontal resistance. Diagnostic markers for *Lr34* would greatly assist wheat breeding programs for this agronomically important resistance gene.

3.3 Wheat genomics in the post-genomic era

The structure of the wheat genome

The genome of bread wheat (*Triticum aestivum* L., genome formula AABBDD) is hexaploid, consists of 16,000 million bases (Mb) and 80% of it is repetitive DNA. This polyploid genome originated naturally from two distinct hybridization events between closely related wild grasses: *Triticum urartu* (A genome), an unidentified relative of *Aegilops speltoides* (B

genome) and *Aegilops tauschii* (D genome) (McFadden and Sears 1946). *Triticum turgidum* (the tetraploid ancestor of durum wheat, genome formula AABB) originated between 0.5 and 3 million years ago. The D-genome was the last to be acquired by polyploid wheat. This occurred about 8,000 years ago and probably only a few times in the first fields of cultivated emmer. Therefore, the level of polymorphism on the D-genome is lower than on the A and B genomes. Besides polyploidy, the large size of each subgenome is due to the high content in transposable elements. These elements are interspersed in the intergenic portion of the genome and they are classified in two main categories: class I elements replicate with a copy-and-past procedure through an mRNA intermediate, elements of class II are moving through a cut and past mechanism which does not conserve their original position (Bennetzen 2000).

Resources for mapping

In order to study the large genome of hexaploid wheat, a number of genetic resources have been created. Already in the 1950ies it was found that viable and genetically stable wheat plants missing one chromosome pair (Sears 1954) or having it substituted by a homologous pair from another genome, could be derived using cytogenetical methods. Later, these lines named 'nullitetrasomic' (Sears 1966a) have been useful to assign certain characters and genes to specific chromosomes before the advent of molecular methodologies (Sears 1966b). More recently, it was observed that certain *Aegilops* chromosomes induce deletions in wheat chromosomes when introduced as monosomic additions (Endo and Gill 1996). Since the wheat genome is highly buffered by polyploidy, many of these deletions are stable and can be maintained in a homozygous state. Over 400 such deletion lines have been produced in wheat (Endo and Gill 1996, Qi et al. 2003) and they have been helpful in anchoring genetic linkage groups to specific chromosomes. The first maps available for wheat were based on RFLP

(Gill et al. 1991, Gale et al. 1995) and later also on SSR markers (Röder et al. 1998; Pestsova et al. 2000, Gupta et a. 2002). This allowed mapping of genes of agronomic importance as well as dissection of the phenotypic variation of quantitative traits (Messmer et al. 2000, Schnurbusch et al. 2004a-b, Paillard et al. 2004). An international effort was established (ITMI, International Triticeae Mapping Initiative) to better characterize the coding portion of the genome, with the development, sequencing and physical mapping of wheat expressed sequence tags (wESTs) to specific bins of the genome. Up to August 2006, more than 500,000 wheat EST sequences have been deposited in the TIGR database (www.tigr.org) and >5,000 of these have been assigned to a specific chromosome bin by deletion mapping (Endo and Gill 1996, Qi et al. 2003).

Resources for gene isolation

While for some plant species important for agriculture or basic research the full genome has been sequenced (e.g. *Arabidopsis thaliana*, *Oryza sativa)* or will soon become available *(Zea mays, Brachypodium distachyon, Vitis vinifera, Medicago truncatula,* etc.), in wheat very little sequence information is available. The size and complexity of the wheat genome and the considerable technical problems in assembling repetitive DNA sequences have so far discouraged the wheat community to undertake a whole genome sequencing initiative. Recently, an international consortium has been established to sequence one of the 21 chromosomes; the wheat chromosome 3B, which is approximately twice the size of the rice genome, has been chosen because it contains many genes of agronomic importance and it was possible to isolate it from the rest of the genome by flow-sorting (Paux et al. 2006). A physical map is in progress for chromosome 3B (Paux et al. 2006) and for the genome of *Ae. tauschii* (Luo et al. 2003), the diploid donor of the hexaploid wheat D-genome, a considerable

number of BAC clones has been assembled into large contigs by fingerprinting. However, up to date there is no complete physical map for any of the wheat chromosomes. The few fully sequenced or partially assembled BAC clones confirmed that most of genomic DNA consists of repetitive elements. Thus, wheat genetic improvement can currently not benefit from the knowledge of its genome sequence.

We are still far from understanding the molecular variation in DNA sequence responsible for the phenotypic variation. Gene characterization can be achieved only after laborious map-based gene isolation (Feuillet et al. 2003, Keller et al. 2005). Some technical advances have recently simplified map-based gene isolation. Arrayed BAC libraries have been developed for wheat diploid ancestors and their close relatives. Subgenomic gridded BAC libraries have been developed using the genotypes of the diploid species *Triticum monococcum* (Lijavetzky et al. 1999), *Triticum urartu, Aegilops speltoides* (Akhunov et. al 2005) and *Aegilops tauschii* (Moullet et. al 1999; Xu et al. 2002; Luo et al. 2003; Akhunov et al. 2005). Specifically, a BAC library from *Ae. tauschii* was developed from the accession 'AL8/78', which has a genome that is very similar to the wheat D-genome. The library consists of more than 300,000 BAC clones and has an 8.5x genome coverage. Around 200,000 clones have been fingerprinted and arranged into ~12,000 contigs (Xu et al. 2002; Luo et al. 2003; http://wheatdb.ucdavis.edu:8080/wheatdb/). This library provides an excellent tool to access large portions of anonymous intergenic sequences from the D-genome. In addition, pooled PCR-screenable BAC libraries have been constructed from the whole genome of hexaploid wheat (Nilmalgoda et al. 2003), as well as from single chromosomes (Safar et al. 2004) and from pooled chromosomes (Janda et al. 2004).

Comparative genomics in cereals

Cereals belong to the grass family, which diverged >50 million years ago. Their genomes vary tremendously in size, chromosome number and ploidy. Differences in genome size are mainly due to the relative amount of repetitive DNA.

	DNA amount (bp)	Genome formula
Rice	$4.4x10^8$	2n = 2x = 24
Sorghum	$7.6x10^8$	2n = 2x = 20
Maize	$2.7x10^9$	2n = 2x = 20
Barley	$5x10^9$	2n = 2x = 14
Wheat	$1.6x10^{10}$	2n = 6x = 42

Despite these differences, gene sequence and order are significantly conserved in cereals. Genomic collinearity at the chromosome level has been revealed by comparative genetic mapping using species-transferable probes, e.g. EST markers. Macrocollinearity was originally summarized by Moore et al. (2005) using what is now known as the 'Circle Diagram'. These studies led to consider grass genomes as a single genetic system. Because of its economic importance and small genome size, rice was the first grass for which a genome sequencing initiative was undertaken. The rice genome was claimed to serve as a 'rosetta stone' for other cereals. Specifically, the rice genome was used as a template to predict gene order and content in all the other cereals with lower gene density, in order to facilitate map-based cloning of genes of interest. In some cases, rice comparative genomics has greatly enhanced development of markers to improve mapping resolution in specific target regions of

larger cereal genomes, e.g. for the isolation of the barley stem rust resistance gene *Rpg1* (Brueggeman et al. 2002). In contrast, in the genomic region of the barley leaf rust resistance *Rph7* a general disruption in collinearity was observed (Brunner et al. 2003). In none of these cases the homologous gene was present in rice. More studies revealed that collinearity was initially overestimated. Mosaic reorganization at the gene level was also observed comparing ~400 kb of maize and sorghum orthologous regions with two rice subspecies (Song et al. 2002). Sequencing information from orthologous regions of several cereals revealed that most of intergenic DNA is not conserved and differences are found both in gene order and content (reviewed by Bennetzen and Ramakrishna 2002, Feuillet and Keller 2002). This is particularly true for resistance genes, which are subject to rapid evolution. The genome sequence of maize, sorghum and of the wild grass *Brachypodium* will soon become available (Wessler et al. 2006, Kresovich et al. 2005, www.brachypodium.org). This will provide genome-wide insights on the extension of microcollinearity conservation and ease identification of new and potentially functional DNA motifs, based on their sequence conservation. Specifically, characterization of small regions from the large Triticeae genomes will greatly profit from comparative genomics with the *Brachypodium* genome. This species is phylogenetically closer to wheat than rice and it will increase the chances of success of comparative genomic approaches.

3.4 Aim of the study

Previous work revealed the presence of a major QTL conferring horizontal leaf rust resistance in the Swiss winter wheat cultivar 'Forno' (Schnurbusch et al. 2004). This QTL showed the same phenotype and genetic mapping location as the durable and widely deployed leaf rust resistance gene *Lr34*. This thesis focuses on the development of new molecular markers to tag

this important source of resistance, using sequence information from the model genomes of the diploid goat grass *(Aegilops tauschii)*, rice *(Oryza sativa)* and the wild grass *Brachypodium sylvaticum.* The new markers developed with this approach revealed dramatic rearrangements between orthologous regions in wheat and rice and enabled us to postulate a candidate gene for *Lr34*. The *Lr34* genomic region was characterized also in the small genome grass species *Brachypodium sylvaticum.* The comparative analysis of the same region in wheat, rice and *Brachypodium* revealed novel aspects of grass genome evolution.

Chapter 4

TAGGING AND VALIDATION OF A MAJOR QTL FOR LEAF RUST RESISTANCE AND LEAF TIP NECROSIS IN THE WINTER WHEAT CULTIVAR 'FORNO'

* Thorsten Schnurbusch, * Eligio Bossolini, Monika Messmer and Beat Keller

** both authors contributed equally to the work*

(2004)

Phytopathology 94 (10): 1036-1041

4. TAGGING AND VALIDATION OF A MAJOR QTL FOR LEAF RUST RESISTANCE AND LEAF TIP NECROSIS IN THE WINTER WHEAT CULTIVAR 'FORNO'

4.1 Abstract

A major leaf rust (*Puccinia triticina* Eriks.) resistance QTL (*QLrP.sfr-7DS*) has previously been described on chromosome 7DS in the winter wheat (*Triticum aestivum* L.) cultivar 'Forno'. It was detected in a population of single seed descent (SSD) lines derived from the cross 'Arina x Forno'. *QLrP.sfr-7DS* conferred a durable and slow-rusting resistance phenotype, co-segregated with a QTL for leaf tip necrosis (LTN) and was mapped close to *Xgwm295* at a very similar location as the adult plant leaf rust resistance gene *Lr34* found in some spring wheat lines. Here, we describe the validation of this QTL by mapping it to the same chromosomal region close to *Xgwm295* on chromosome 7DS in a population of SSD lines from the winter wheat x spelt (*T. spelta* L.) cross 'Forno x Oberkulmer'. In both populations, the LOD curves for leaf rust resistance and LTN peaked at identical or very similar locations, indicating that both traits are due to the same gene. We have improved the genetic map in the target region of *QLrP.sfr-7DS* using microsatellite and EST markers. Two EST loci (*Xsfr.BF473324* and *Xsfr.BE493812*) define a genetic interval of 7.6 cM containing *QLrP.sfr-7DS*, a considerably more precise genetic location for this QTL than previously described both in spring and winter wheat. The identified genetic interval is physically located in the distal thirty-nine percent of chromosome 7DS. Single marker analysis identified *Xsfr.BF473324* and *Xgwm1220* as the most informative loci for *QLrP.sfr-7DS* and *QLtn.sfr-7DS*. In the rice genome, the two ESTs flanking the *QLrP.sfr-7DS/QLtn.sfr-7DS*

chromosomal segment in wheat are conserved on chromosome 6S in a region colinear with wheat chromosome 7DS. There, they define a physical region of three rice BACs spanning ~ 300 kb.

4.2 Introduction

From the fifty known leaf rust (*Puccinia triticina* Eriks.) resistance genes, two of them, *Lr34* and *Lr46*, have been classified as slow-rusting genes (van der Plank 1963, Parlevliet 1979, Singh and Gupta 1991, Singh and Rajaram 1992). The origin of *Lr34* can be traced back to the Brazilian wheat (*Triticum aestivum* L., 2n = 6x = 42) *cv*. 'Frontana' which was already released in 1943 (Singh and Rajaram 1992). In contrast to many race-specific resistance genes, *Lr34* proved to confer durable resistance over a long period of time, in different environments as well as against diverse pathotypes of the fungus. Genetic studies using lines from the CIMMYT (Singh 1993, Nelson et al. 1997), the Japanese (Suenaga et al. 2003) or the Indian (Kaur et al. 2000) spring wheat germplasm confirmed the involvement of the gene in the expression of durable leaf rust resistance. Recently, *Lr34* or a winter wheat allele of it, *QLrP.sfr-7DS*, was also detected in the Swiss winter wheat *cv*. 'Forno' which has shown durable field resistance against leaf rust (Schnurbusch et al. 2004).

During the seedling stage, wheat lines with *Lr34* show a susceptible infection type (Rubiales and Niks 1995). In the adult plant stage, the *Lr34* phenotype typically exhibits a rust pustule gradient on the flag leaf with more and larger pustules at the leaf basis and less and smaller pustules towards the leaf tip. Usually, the observation of a pustule gradient goes along with the presence of leaf tip necrosis (LTN) (Singh 1992a,b). From the CIMMYT spring wheat germplasm, it is known that the leaf tip necrosis (*Ltn*) gene conferring LTN is associated with

the slow-rusting and durable adult plant resistance gene *Lr34* (Singh 1992a). However, it is still not yet clear whether *Ltn* is identical to *Lr34* but there are strong indications for the pleiotropic action of a single gene (Messmer et al. 2000, Schnurbusch et al. 2004). In addition, the *Ltn*/*Lr34* chromosomal segment also confers a durable resistance to stripe (yellow) rust, *Yr18* (Singh et al. 1992b), stem rust (Dyck 1987, Liu and Kolmer 1998), as well as tolerance against the barley yellow dwarf virus, *Bdv1* (Singh 1993, Ayala et al. 2002) and inactivates a stem rust resistance suppressor (Kerber and Aung 1999). Therefore, the *Ltn*/*Lr34* chromosomal segment is unique and a very valuable source of broad-spectrum and durable disease resistance in wheat breeding.

Lr34 has been mapped to the short arm of chromosome 7D using cytogenetic analysis (Dyck 1987). Later, mapping studies revealed that *Lr34* is close (two to five cM) to the microsatellite marker GWM295 (Schnurbusch et al. 2004, Suenaga et al. 2003) and the RFLP probe WG834 (Nelson et al. 1997) on chromosome arm 7DS. So far, there is no tightly linked (< 1 cM) molecular marker available for *Lr34* or its winter wheat allele which could be used for marker-assisted selection (MAS). The sequence information of the rice genome (Feng et al. 2002, Goff et al. 2002, Sasaki et al. 2002) with its close relatedness and partial colinearity to wheat (Gale and Devos 1998, Sorrells et al. 2003) enables efficient generation of new probes for genetic mapping in wheat. In addition, almost 10,000 wheat expressed sequence tags (ESTs) have been physically mapped using bin-mapping and chromosome deletion lines (Endo and Gill 1996, Qi et al. 2003). This physical mapping information is very helpful to derive markers for specific chromosomal regions in wheat. In the last few years, there has been intensive efforts in wheat genomics with the goal to isolate agronomically important genes. Recently, it was shown that chromosomal walking in wheat is feasible (Stein et al. 2000, Faris et al. 2003). In addition, two major disease resistance genes against leaf rust, *Lr10*

(Feuillet et al. 2003) and *Lr21* (Huang et al. 2003), have been isolated. So far, no durable or slow-rusting leaf rust resistance gene has been isolated in wheat.

In this study, we could narrow down the genetic interval containing the putative winter wheat allele of *Lr34* in *cv*. 'Forno' to a region of 7.6 cM on chromosome 7DS. The flanking markers in wheat are colinear in the rice genome, defining a genomic region of ~300 kb.

Materials and methods

Linkage analysis and physical mapping

Linkage analysis was carried out using the computer package MAPMAKER 3.0b for MS-DOS (Lander et al. 1987). All new simple sequence repeat markers (SSR), ESTs or restriction fragment length polymorphism (RFLP) markers in the 'Arina' x 'Forno' population were mapped using 240 single seed descent (SSD)-lines as described in (Paillard et al. 2003). In the 'Forno' x 'Oberkulmer' population, mapping was accomplished according to (Messmer et al. 2000) and based on 204 lines. The establishment of the new linkage group on the short arm of chromosome 7D in the 'Forno' x 'Oberkulmer' population was conducted using the procedures described in (Paillard et al. 2003). For adding new markers to chromosome 7D in the 'Arina' x 'Forno' population, we used the 'try' command and validated the new marker sequence by using the 'ripple' command. Linkage groups were drawn with the software package MapChart (Voorrips 2002). Physical mapping and bin assignment of fifteen markers was conducted using the deletion line 7DS-4 (breakpoint 0.61) derived from *cv*. 'Chinese

Spring' in which thirty-nine percent of the short arm of chromosome 7D was missing (Endo and Gill 1996, Qi et al. 2003).

QTL mapping

Interval QTL analysis was carried out with the composite interval mapping (CIM) program PLABQTL, Version 1.1 (Utz and Melchinger 2000) which is based on multiple regression. Model and covariate selection as well as determination of a significant LOD threshold were applied as described in (Schnurbusch et al. 2004). The phenotypic data for the QTL analysis in the 'Arina' x 'Forno' population were taken from the estimated genotypic values across seven environments as the area under the disease progress curve (AUDPC_%) for the percentage of infected leaf area, whereas the means of leaf tip necrosis (LTN) were obtained by measuring in six field environments (Schnurbusch et al. 2004). In 'Forno' x 'Oberkulmer', scoring across three field environments on a scale from one (fully resistant) to nine (fully susceptible) was used to estimate the phenotypic means of the population for leaf rust resistance, whereas LTN was quantitatively measured as the length of the necrotic leaf tip (Messmer et al. 2000). Single marker analysis was performed by a simple one-way ANOVA using the SAS GLM procedure (SAS Institute 1991).

Sequence analysis and data sources

In order to identify rice BAC clones with nucleotide sequence similarity to the wheat ESTs assigned to physical bin 7DS-4, we used a local BLAST server to performed BLASTn searches (Altschul et al. 1997). The sequence of wheat ESTs were used as queries against

local databases composed of the rice genome and the BAC clones assigned to rice chromosome 6. The sequences of *O. sativa* ssp. *japonica* BACs were downloaded from the RiceGAAS web site (ftp://ftp.dna.affrc.go.jp/pub/RiceGAAS), as of May 2003 and the relative positions of BACs within the rice chromosome 6 were estimated using the IRGSP mapping data (http://rgp.dna.affrc.go.jp/IRGSP). Triticeae EST sequences anchored within chromosome 7DS and assigned by chromosome bin mapping (Qi et al. 2003) were downloaded from the GrainGenes web site (http://wheat.pw.usda.gov/NSF/progress_mapping.html). BLASTn results were parsed: a score of 50 or higher were set as threshold limits and the redundancies due to overlapping BAC clones were removed from raw data. Detailed sequence analysis were performed with the EMBOSS package (Rice et al. 2000).

4.4 Results

Genetic and physical mapping of wheat chromosome 7DS

Taking the map of chromosome 7D in the 'Arina' x 'Forno' population as a starting point (Paillard et al. 2003), eleven new markers were added to this chromosome (Fig. 1). The entire linkage group now spans 280.5 cM. From the eleven additional markers, seven were PCR-based SSRs, three were ESTs and one was an RFLP marker (CDO475). For many of these markers only their chromosomal or physical bin location was previously known.

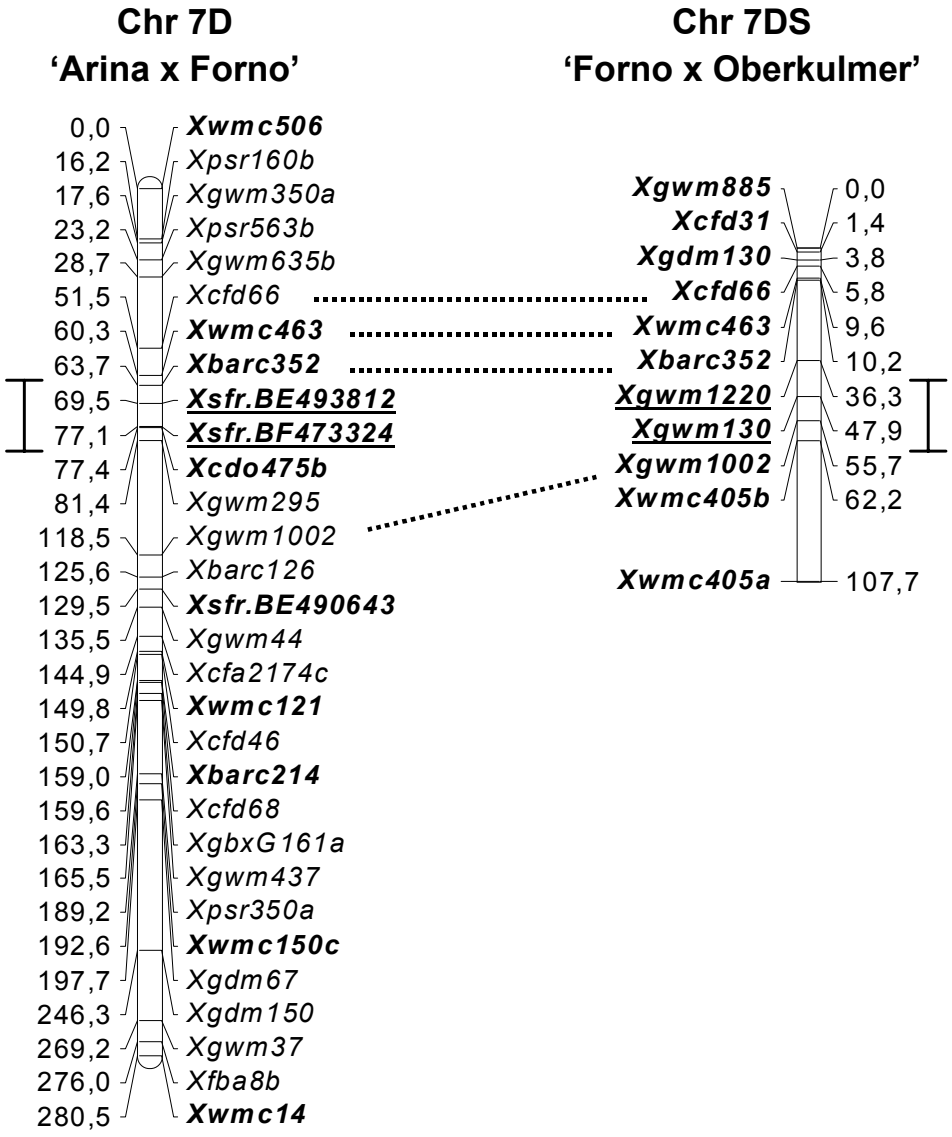

Figure 1. Genetic linkage maps of wheat chromosome 7D for the 'Arina x Forno' and 'Forno x Oberkulmer' populations. Loci in bold represent newly added markers compared to the previously published linkage maps (Messmer et al. 1999, Paillard et al. 2003). Black bars beside the two linkage groups and underlined loci indicate the most probable marker interval of *QLrP.sfr-7DS*/*QLr.sfr-7DS* whereas dotted lines show common markers for both populations.

Due to lacking polymorphism on the short arm of chromosome 7D using only RFLP markers, this chromosomal region was not covered in the published 'Forno' x 'Oberkulmer' map (Messmer et al. 1999). With the recent characterization of numerous wheat SSR markers, it

was possible to establish a genetic linkage map for this chromosomal region in the 'Forno' x 'Oberkulmer' population (Fig. 1). Ten out of twenty-four SSRs tested were polymorphic and formed a linkage group of 107.7 cM. Four SSRs were mapped in both populations with the same relative marker order (Fig. 1).

Physical mapping of fifteen markers revealed that fourteen markers failed to hybridize to or amplify from DNA of the 7DS-4 deletion line when compared with amplifications or hybridizations of DNA from *cv.* 'Chinese Spring' indicating that the 7DS-4 breakpoint 0.61 is distal to *Xcfd46* (Fig. 2). Two SSR markers, BARC352 and BARC154 (the latter is monomorphic in both populations), revealed two loci being located in bin 7DS-4. Bin mapping located *Xgwm295*, previously the closest marker proximal to *Lr34* (Suenaga et al. 2003) and *QLrP.sfr-7DS* (Schnurbusch et al. 2004), being located in the 7DS4-0.61-1.00 physical bin of chromosome 7DS (Fig. 2). Therefore, this physical region also contains *QLrP.sfr-7DS*.

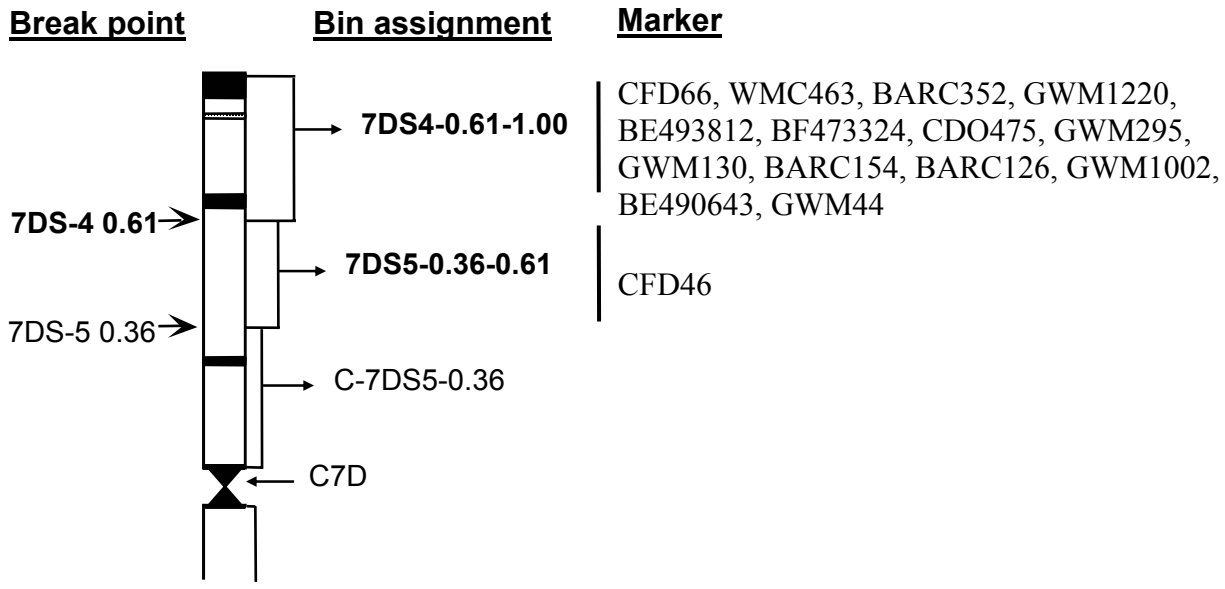

Figure 2. Physical bin-mapping of molecular markers using the deletion line 7DS-4 (breakpoint 0.61). Markers which failed to hybridize to or amplify from DNA of this deletion line, but hybridized to or amplified from DNA of *cv.* 'Chinese Spring', were assigned to the chromosomal bin 7DS4-0.61-1.00.

Identification of an orthologous chromosomal region in rice close to QLrP.sfr.7DS and its use to derive new markers

Southern blot analysis of a putative inorganic vacuolar pyrophosphatase (wheat EST accession numbers BE427461, BE490643) revealed a single copy of this gene in each subgenome of hexaploid wheat. It was already mapped to chromosome 7BS, *Xsfr.BE427461*, in the 'Arina x Forno' population (Paillard et al. 2003). The wheat EST BE490643 could be mapped approximately 50 cM proximal to *Xgwm295* to a homoeologous locus on the short arm of chromosome 7D (Fig. 1) using a different set of restriction enzymes cleaving DNA from *cv*. 'Chinese Spring' and the 7DS-4 deletion line in order to identify the 7D polymorphism for southern mapping. Blast analysis (BLASTn) of this wheat EST in rice revealed two loci on rice chromosome 6.

Using the homoeologous relationships between group 7 chromosomes in wheat (Nelson et al. 1995), the approximate map position for *Xcdo475b* was estimated being close to *Lr34/QLrP.sfr-7DS*. *Xcdo475b* mapped 4 cM distal to *Xgwm295* in the 'Arina x Forno' population (Fig. 1) and was closer to *QLrP.sfr-7DS* than *Xgwm295* (Fig. 3A and 4). Sequencing of the anonymous cDNA probe CDO475 from oat (*Avena sativa* L.) revealed a putative carboxypeptidase. Its sequence gave a single BLASTn match on the short arm of rice chromosome 6S. The rice ortholog of CDO475 was found on rice BAC AP003487 (Fig. 4).

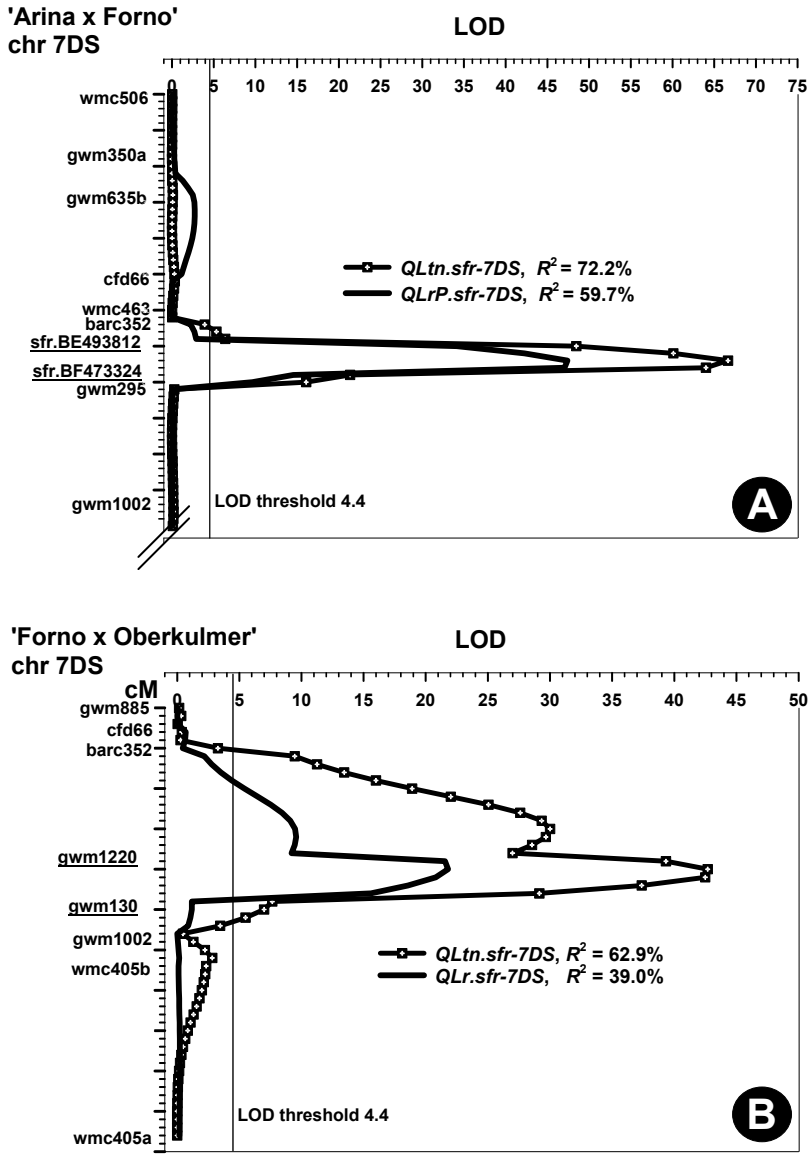

Figure 3. LOD curves for leaf rust resistance and LTN on chromosome 7DS. (3A) population 'Arina x Forno' (3B) population 'Forno x Oberkulmer'. Underlined loci indicate the most probable marker interval of QLrP.sfr-7DS /QLr.sfr-7DS. One dash on the vertical axis corresponds to two cM.

In March 2003, 201 loci corresponding to 164 wheat ESTs were physically assigned to bin 7DS-4 (http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus.cgi) which is partially orthologous to rice chromosome 6 (Gale and Devos 1998, Sorrells et al. 2003). We have blasted (BLASTn) the 164 wheat ESTs against rice chromosome 6, and blast results with a cut off score value below 50 were discarded. The remaining wheat ESTs were physically

arranged according to blast analysis on BACs/PACs of rice chromosome 6. Wheat ESTs that had rice orthologs closely linked to the rice cdo475 gene (carboxypeptidase), were selected based on low gene copy number in their southern blot analysis (http://wheat.pw.usda.gov/cgi-bin/westsql /map_locus.cgi). The rice ortholog of the wheat EST BF473324 (putative GTP-binding membrane protein) was directly adjacent to the rice carboxypeptidase gene on the same rice BAC AP003487 (Fig. 4). In wheat, cdo475 and EST BF473324 were genetically tightly linked (0.3 cM, Fig. 1). The wheat EST BE493812 (hypothetical protein) represents the distal end of the *QLrP.sfr-7DS* and *QLtn.sfr-7DS* interval in wheat (see below). Its sequence blasted to BAC AP003708 in the same region on rice chromosome 6 (Fig. 4). Thus, the two loci, *Xsfr.BF473324* and *Xsfr.BE493812*, delimiting *QLrP.sfr-7DS* and *QLtn.sfr-7DS* in a 7.6 cM interval in hexaploid wheat (see below) correspond to approximately three rice BACs or ~300 kb in the orthologous region in rice (Fig. 4).

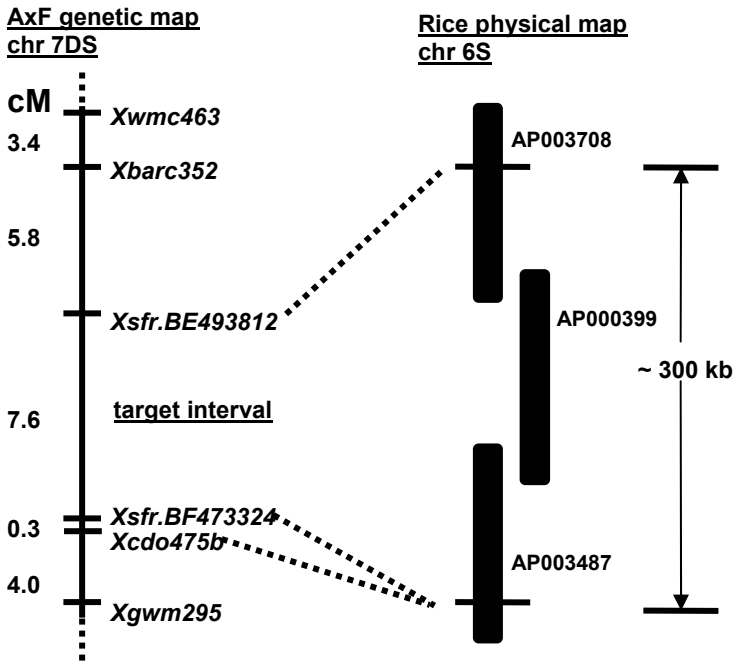

Figure 4. Genetic and physical relationships between rice chromosome 6S and wheat chromosome 7DS. The target interval in the genetic map of 'Arina x Forno' (7.6 cM) of *QLrP.sfr-7DS* was flanked by two wheat EST loci, *Xsfr.BE493812* and *Xsfr.BF473324*. Their orthologous loci in rice were detected on BACs/PACs AP003708 and AP003487 which flank a chromosomal region of approximately 300 kb on rice chromosome 6S.

Mapping QTL for leaf rust resistance and leaf tip necrosis in the improved genetic map of chromosome 7DS

QTL mapping (composite interval mapping (CIM) and LOD>4.4) using the improved map of chromosome 7DS in the 'Arina' x 'Forno' population resulted in the definition of a small QTL interval (Fig. 3A) for leaf rust resistance (*QLrP.sfr-7DS*) and LTN (*QLtn.sfr-7DS*). Both QTL were flanked by *Xsfr.BE493812* and *Xsfr.BF473324*, reducing the marker interval from 60 cM (Schnurbusch et al. 2004) to 7.6 cM. Still, both LOD curves were completely overlapping and peaked 3.1 cM (± 2 cM for 1-LOD-support-interval) distal to *Xsfr.BF473324* indicating pleiotropism between *QLrP.sfr-7DS* and *QLtn.sfr-7DS* (Messmer et al. 2000, Schnurbusch et al. 2004). Due to the closer flanking markers, the percentage of the phenotypic variation explained increased to 59.7% with a corresponding LOD value of 47.4 for *QLrP.sfr-7DS* and to 72.2% with a LOD value of 66.7 for *QLtn.sfr-7DS*. Single marker analysis for this chromosomal region identified *Xsfr.BF473324* as the most informative locus (Table 1) giving a coefficient of determination (R^2 in percent) of 29.4 for leaf rust resistance and 61.8 for LTN, respectively. The marker locus *Xcdo475b* was tightly linked to *Xsfr.BF473324* and therefore, it also explained a high percentage of the observed phenotypic variation for both traits within the 'Arina' x 'Forno' population (Table 1).

QTL analysis of the linkage group representing chromosome 7DS in the 'Forno' x 'Oberkulmer' population revealed a major QTL for leaf rust resistance (designated as *QLr.sfr-7DS*) and LTN (Fig. 3B). Also in the 'Forno x Oberkulmer' population, the two LOD curves for leaf rust resistance and LTN largely overlapped and were very similar, but did not peak at the identical position. *QLr.sfr-7DS* was located 2.7 cM (± 3 cM) proximal to *Xgwm1220*, whereas *QLtn.sfr-7DS* was 4.7 cM (± 3 cM) proximal to *Xgwm1220*. *QLr.sfr-7DS* accounted for 39% of the observed phenotypic variation for leaf rust resistance in the 'Forno' x

'Oberkulmer' population with a corresponding LOD value of 21.8 (Fig. 3B). *QLtn.sfr-7DS* reached a LOD value of 43.2 explaining 62.8% of the phenotypic variance for LTN. Both QTL were most likely localized within the *Xgwm1220-Xgwm130* interval (Fig. 3B). Single marker analysis of this linkage group in the 'Forno x Oberkulmer' population supported this hypothesis by identifying *Xgwm1220* as the best (R^2 of 22.7% for leaf rust resistance and 65.4% for LTN, respectively) and *Xgwm130* as the second best (21.9% and 46%, respectively) locus within this linkage group (Table 1).

Table 1. Single marker analysis for leaf rust resistance and leaf tip necrosis (*Ltn*) on the short arm of chromosome 7D. Coefficients of determination (R^2) in percent are given for the 'Arina' x 'Forno' as well as the 'Forno' x 'Oberkulmer' population and loci closely mapping to *QLrP.sfr-7DS* or *QLr.sfr-7DS*, respectively, and *QLtn.sfr-7DS*.

Locus	'Arina' x 'Forno'		'Forno' x 'Oberkulmer'	
	leaf rust resistance R^2 (%)	leaf tip necrosis R^2 (%)	leaf rust resistance R^2 (%)	leaf tip necrosis R^2 (%)
Xcfd66	5.2	20.1	6.9	22.9
Xwmc463	11.5	29.1	2.7	19.6
Xbarc352	13.0	35.7	2.7	20.3
Xsfr.BE493812	18.7	46.5	-	-
Xgwm1220	-	-	22.7	65.4
Xsfr.BF473324	29.4	61.8	-	-
Xcdo475b	27.6	59.4	-	-
Xgwm295	21.9	47.7	-	-
Xgwm130	-	-	21.9	46.0
Xgwm1002	3.1	4.1	17.6	16.8
Xwmc405b	-	-	14.2	7.2

4.5 Discussion

Tagging and validation of QLrP.sfr-7DS and QLtn.sfr-7DS

The major leaf rust resistance QTL of *cv.* 'Forno', *QLrP.sfr-7DS*, was physically mapped to the distal thirty-nine percent of the short arm of chromosome 7D. In the 'Arina' x 'Forno' population, it was flanked by two wheat ESTs defining a 7.6 cM interval. Now, *QLrP.sfr-7DS* was genetically localized approximately 7.0 cM distal to *Xgwm295* in a considerably more precise interval for this QTL than previously described both in spring (Suenaga et al. 2003) and winter wheat (Schnurbusch et al. 2004). The contributions of *QLrP.sfr-7DS* to leaf rust resistance was validated in another mapping population of *cv.* 'Forno' and a winter spelt (*cv*. 'Oberkulmer'). The map location of *QLrP.sfr-7DS* and *QLr.sfr-7DS* in the winter wheat *cv*. 'Forno' is very similar to previously reported mapping data for the durable resistance gene *Lr34* in several spring wheat lines (Nelson et al. 1997, Suenaga et al. 2003). The *Lr34* gene acts as a quantitative gene conferring durable leaf rust resistance, particularly in combination with other major resistance genes. Its activity has been validated in several populations derived from different spring wheat lines (German and Kolmer 1992, Sawhney 1992, Kolmer 1996, Klopper and Pretorious 1997, Kaur et al. 2000). We have found a complete linkage between *QLrP.sfr-7DS* and leaf tip necrosis. Similarly, it was found that the slow-rusting phenotype of *Lr34* and leaf tip necrosis were co-segregating in spring wheat lines carrying *Lr34* (Singh 1992a).

Molecular analysis of *QLrP.sfr-7DS*/*QLtn.sfr-7DS* showed that it was genetically localized in the same chromosomal region as *Lr34* (Nelson et al. 1997, Suenaga et al. 2003). Phenotypically, *QLrP.sfr-7DS*/*QLtn.sfr-7DS* was also very similar to *Lr34*, because we found

lines in our mapping populations exhibiting the typical "*Lr34*-phenotype" during the adult plant stage in the field (Schnurbusch et al. 2004). There, flag leaves of infected lines in the field showed a rust pustule gradient on the leaf with more and larger pustules at the leaf basis and less and smaller pustules towards the leaf tip. Usually, the pustule gradient went along with the presence of LTN (Singh 1992a, Kolmer 1996). However, the *Lr34* gene was initially defined within the CIMMYT spring wheat germplasm, as it was derived from *cv*. 'Frontana' (Singh et al. 1998). To date, there is no pedigree link between *cv*. 'Forno' and *cv*. 'Frontana' indicating either another source for *Lr34* different from *cv*. 'Frontana' or the occurrence of an independent winter wheat allele of the *Lr34* gene within the winter wheat germplasm.

Merging the mapping data from both Swiss winter wheat populations, one can conclude that the position of the *QLrP.sfr-7DS* chromosomal segment is most likely between *Xgwm1220* and *Xsfr.BF473324*. Thus, the SSR marker GWM1220, approximately two to three cM away from *QLrP.sfr-7DS*, is the closest PCR-based marker currently available. However, GWM1220 was not polymorphic in the 'Arina' x 'Forno' population and therefore, cannot be considered as a diagnostic marker. Nevertheless using SSR and wheat EST markers, we substantially increased the marker density on chromosome 7DS in both mapping populations although the low level of polymorphism particularly for RFLP and EST markers on the D genome of hexaploid wheat has been well documented (Cadalen et al. 1997, Messmer et al. 1999, Paillard et al. 2003).

Colinearity between rice chromosome 6S and wheat chromosome 7DS

The rice genome sequence constitutes a model for other genomes of many related grass species such as maize (*Zea maize* L.), sorghum (*Sorghum* sp. L.), barley (*Hordeum vulgare* L.)

or wheat (Laurie and Devos 2002, Sorrells et al. 2003). We compared rice chromosome 6S and wheat chromosome 7DS at four orthologous loci and found colinearity of gene content and gene order in the region described. In wheat and rice, the putative pyrophosphatase gene (*Xsfr.BE490643*) mapped proximal to the target interval flanked by *Xsfr.BE493812* (hypothetical protein) and the putative GTP-binding membrane protein (*Xsfr.BF473324*). The orientation of the identified target interval containing *QLrP.sfr-7DS* was also conserved in both species, as *Xsfr.BE493812* mapped distal to the gene encoding a GTP-binding membrane protein, which itself was tightly linked to the putative carboxypeptidase (*Xcdo475b*). Despite the huge differences between the genomes of rice and wheat in chromosome number, ploidy level, and size, they often show a very good conservation of marker order (colinearity). Early reports found high colinearity between rice and the Triticeae (wheat, barley, oat, rye) species (Kurata et al. 1994, VanDeynze et al. 1995, Gale and Devos 1998, Gallego et al. 1998, Han et al. 1998). Detailed studies at the micro-level revealed rearrangements of gene content displaying a mosaic organization of the orthologous regions with conserved sequences interspersed with non-conserved sequences (Foote et al. 1997, Keller and Feuillet 2000, Bennetzen and Ramakrishna 2002, Song et al. 2002, Brunner et al. 2003, Guyot et al. 2003). However, during the map-based cloning of the *Rpg1* stem rust resistance gene in barley excellent colinearity between the distal end of the short arm of rice chromosome 6 and barley chromosome 7HS was demonstrated (Brueggeman et al. 2002, Horvath et al. 2003).

Therefore, in order to "mendelize" a quantitative trait/resistance into a single locus, rice sequence information in conjunction with mapped wheat ESTs will help to dissect the present target interval for fine-mapping. In addition, newly released BAC libraries from *Aegilops tauschii*, the D-genome donor of hexaploid wheat (http://wheat.pw.usda.gov/PhysicalMapping/progress.html), will also be very valuable tools to conduct high-resolution mapping in wheat. The degree of conserved colinearity between rice chromosome 6S and wheat

chromosome 7DS will essentially determine whether fine-mapping and cloning of *QLrP.sfr-7DS* will be rapidly possible.

Outlook

So far, despite enormous economic interest no durable or slow-rusting leaf rust resistance gene has been isolated in wheat. The isolation and cloning of *QLrP.sfr-7DS* in wheat will be the first step of a physiological understanding of slow-rusting or partial disease resistance and will also clarify whether these slow-rusting genes are "defeated" major resistance genes (Chantret et al. 1999, Li et al. 1999). *QLrP.sfr-7DS* seems to be a very suitable candidate for a map-based cloning strategy of a durable and broad-spectrum disease resistance gene in hexaploid wheat.

Acknowledgements:

We are very thankful to Romain Guyot for his bioinformatical support and to Dr. H.F. Utz (University of Hohenheim, Germany) for providing the computer package PLABQTL. We thank Dr. P. Cregan and Dr. R. Ward (Beltsville Agricultural Research Center, MD, USA) for publicly available primer sequence and map location of microsatellite markers (barc). We thank Dr. Olin Anderson and Dr. Curt Crossman (USDA-ARS, Albany, CA, USA) for wheat ESTs, Dr. Mark Sorrells (Cornell University, Ithaca, NY, USA) for cdo475 and Dr. Marion Röder (IPK Gatersleben, Germany) as well as TraitGenetics Inc. (Gatersleben, Germany) for providing unpublished primer sequences. We are grateful to Dr. Jon Raupp (KSU, Manhattan, KS, USA) for providing the seeds of the deletion line CS7DS-4. This work was supported by the University of Zürich and the Swiss National Science Foundation (SNF grant 3100-065114). The authors declare that the experiments conducted for this publication comply with the current laws of Switzerland.

Chapter 5

DEVELOPMENT OF SSR MARKERS SPECIFIC FOR THE *Lr34* RESISTANCE REGION OF WHEAT USING SEQUENCE INFORMATION FROM RICE AND *Aegilops tauschii*

* Eligio Bossolini, * Simon G. Krattinger and Beat Keller

* *both authors contributed equally to the work*

(2006)

Theoretical and Applied Genetics, in press

DOI: 10.1007/s00122-006-0364-5

online publication August 8th 2006

5. DEVELOPMENT OF SSR MARKERS SPECIFIC FOR THE *Lr34* RESISTANCE REGION OF WHEAT USING SEQUENCE INFORMATION FROM RICE AND *Aegilops tauschii*

5.1 Abstract

Hexaploid wheat (*Triticum aestivum* L.) originated about 8,000 years ago from the hybridization of tetraploid wheat with diploid *Aegilops tauschii* Coss. containing the D-genome. Thus, the bread wheat D-genome is evolutionary young and shows a low degree of polymorphism in the bread wheat gene pool. To increase marker density around the durable leaf rust resistance gene *Lr34* located on chromosome 7DS, we used molecular information from the orthologous region in rice. Wheat expressed sequence tags (wESTs) were identified by homology with the rice genes in the interval of interest, but were monomorphic in the 'Arina' x 'Forno' mapping population. To derive new polymorphic markers, bacterial artificial chromosome (BAC) clones representing a total physical size of ~1 Mb and belonging to four contigs were isolated from *Ae. tauschii* by hybridization screening with wheat ESTs. Several BAC clones were low-pass sequenced, resulting in a total of ~560 kb of sequence. Ten microsatellite sequences were found, and three of them were polymorphic in our population and were genetically mapped close to *Lr34*. Comparative analysis of marker order revealed a large inversion between the rice genome and the wheat D-genome. The SWM10 microsatellite is closely linked to *Lr34* and has the same allele in the three independent sources of *Lr34:* 'Frontana', 'Chinese Spring' and 'Forno', as well in most of the

genotypes containing *Lr34*. Therefore, SWM10 is a highly useful marker to assist selection for *Lr34* in breeding programs worldwide.

5.2 Introduction

Bread wheat (*Triticum aestivum* L., genome formula AABBDD) is a crop with one of the largest known genomes and a low gene density. Genetically, wheat is an allohexaploid species that evolved by hybridization of three closely related wild grasses: *Triticum urartu* (A genome), a relative of *Aegilops speltoides* (B genome) and *Aegilops tauschii* (D genome) (McFadden and Sears 1946). Among the three wheat genomes, the D-genome was the last to be acquired by polyploid wheat, about 8,000 years ago. Although the hybridization with *Ae. tauschii* seems to be polyphyletic, the D-genome has a low degree of polymorphism within the cultivated germplasm (reviewed by Ogbonnaya et al. 2005).

In wheat, simple sequence repeat markers (SSRs) are a powerful tool for genetic mapping (Röder et al. 1998; Pestsova et al. 2000), diversity analysis (Huang et al. 2002) and marker assisted selection (Dubcovsky 2004). Wheat SSRs have been mostly developed on a genome-wide scale (Röder et al. 1998; Gupta et al. 2002; Song et al. 2005). Consensus mapping of publicly available microsatellites has provided an overview of the saturation reached in the hexaploid wheat genome (Somers et al. 2004). However, with a total genome size of 16,000 Mbp and a content in repetitive DNA estimated to be ~80%, there is little chance that a randomly developed marker tags a small and specific genomic region of interest. In the past, the development of microsatellite markers for specific regions of the wheat genome was not possible because of technical limitations and lack of genomic resources. More recently, some technical advances have allowed to partially overcome these problems (reviewed by Keller et

al. 2005). Arrayed BAC libraries have been developed for wheat diploid ancestors and their close relatives. Specifically, a BAC library from *Ae. tauschii* was developed from the accession 'AL8/78' which has a genome that is closely related to the wheat D-genome. The library consists of more than 300,000 BAC clones and has a 8.5x genome coverage. Around 200,000 clones have been fingerprinted and arranged into ~12,000 contigs (Luo et al. 2003; http://wheatdb.ucdavis.edu:8080/wheatdb/). This library provides an excellent tool to access large portions of anonymous intergenic sequences from the D-genome. In addition, pooled PCR-screenable BAC libraries have been constructed from genomic DNA of hexaploid wheat (Nilmalgoda et al. 2003).

Since the sequence of the rice genome has been released (assembly 4 of TIGR), a thorough comparative analysis of the rice and wheat genomes has demonstrated that, based on genome synteny, it is possible to predict gene order in wheat using the rice sequence as template (Sorrells et al. 2003). However, numerous exceptions to collinearity have also been found between rice and the family of the Triticeae, complicating approaches relying on comparative genomics (Bennetzen and Ramakrishna 2002; Brunner et al. 2003; Guyot et al. 2004).

The combined use of new molecular tools in wheat and rice has a high and unexploited potential for the development of microsatellite markers tagging small chromosomal regions in bread wheat. The isolation of microsatellite markers from BAC libraries for targeted genomic regions was described first in soybean by Cregan et al. (1999) and later in several other crops. Recently it was also reported in wheat by Shen et al. (2005) using an arrayed BAC library of cv. 'Chinese Spring', and the markers developed were tagging loci from all the three genomes. The close similarity of the bread wheat D-genome to the genome of *Ae. tauschii* provides a unique opportunity to use genomic information derived from *Ae. tauschii* for marker

development in wheat. High transferability of SSRs from *Ae. tauschii* to wheat was demonstrated by Guyomarc'h et al. (2002).

One of the most relevant genes in wheat disease resistance breeding is *Lr34,* located on chromosome 7DS (Suenaga et al. 2003; Schnurbusch et al. 2004a,b). *Lr34* confers adult plant, durable resistance to leaf rust in bread wheat, and it was originally detected in the spring wheat material of CIMMYT (Singh 1992). In the Swiss winter wheat cv. 'Forno', we identified a quantitative trait locus (QTL) for leaf rust resistance, *QLrP.sfr-7DS,* which had a very similar mapping location and phenotype as *Lr34* (Schnurbusch et al. 2004a). Therefore, although no records are available to link the pedigree of 'Forno' to the spring wheat germplasm at CIMMYT, *QLrP.sfr-7DS* and *Lr34* are possibly the same locus. *Lr34* is associated with the morphological character 'leaf tip necrosis' (LTN, Singh 1992) and the stripe rust resistance gene *Yr18* (McIntosh 1992). The *Lr34* locus has provided durable resistance to both leaf rust (causal fungus: *Puccinia triticina* Eriks.) and stripe rust (*Puccinia striiformis* Westend. *f. sp. tritici,* Ma and Singh 1996), making *Lr34* a unique resource for breeding and a model for understanding the molecular basis of partial horizontal resistance. Previous work in the 7DS chromosomal region demonstrated close genetic linkage of *Lr34* with SSRs GWM130, GWM295, GWM1220, and ESTs CDO475, BF473324 and BE493812 (Suenaga et al. 2003, Schnurbusch et al. 2004b, Spielmeyer et al. 2005). However, there is still no diagnostic marker available.

The aim of this study was to develop new microsatellite markers specific for the chromosomal region of *Lr34,* based on orthologous regions in rice and *Ae. tauschii*. In addition, we wanted to analyze the newly developed markers for allelic diversity in a set of wheat lines with or without *Lr34* and to test their diagnostic value for marker-assisted selection.

5.3 Materials and methods

Plant material

Two *Lr34* recombinant populations were used for genetic mapping: 240 'Arina' x 'Forno' $F_{5:7}$ recombinant inbred lines (Paillard et al. 2003) and a set of 600 F2 plants derived from a cross between 'Arina' and an 'Arina-*Lr34'* introgression line (BC2F4, with 'Arina' as the recurrent parent). This introgression line was verified to be homozygous for the 'Forno' *Lr34* segment by phenotypic evaluation and with the *Lr34* flanking markers *Xsfr.BE493812/PstI* and *Xswm5* (developed within this study).

An allele diversity study was performed on a set of lines from different breeding programs around the world for which information was available on *Lr34*. This set contains five 'Chinese Spring' chromosome-substitution lines, having the chromosome 7D of 'Chinese Spring' (which has *Lr34*, Dyck 1991) replaced by chromosome 7D of lines which do not have *Lr34* ('Cheyenne', 'Hope', 'Red Egyptian', 'Timstein' and 'Thatcher'); the line 'Lalbahadur' (without *Lr34*) and the corresponding 7D substitution line 'Lalbahadur-Parula 7D' ('Parula' has *Lr34,* Singh 1992); the near isogenic lines 'Jupateco 73' and 'Jupateco 73R' (with *Lr34,* Singh 1992); 'Avocet' and 'Avocet R' (with *Lr34*); 'Thatcher *Lr16*' and 'Thatcher *Lr16* + *Lr34*'.

Development of probes from the Lr34 *orthologous region in rice*

The rice BAC clones orthologous to the wheat '*Lr34* orthologous region' defined by the two flanking markers *Xsfr.BE493812* and *Xsfr.BF473324* were identified by Schnurbusch et al. (2004b). The nucleotide sequences of genes predicted by RICEGAAS (Sakata et al. 2002) in the rice BAC clones AP000399, AP003708, AP003487, AP003767 and AP003632 were compared with BLASTn to Triticeae and oat ESTs. A set of 24 ESTs of wheat, barley and oat was identified with a similarity threshold value of e^{-15} and was hybridized to wheat Southern blots to test for copy number. Sixteen probes had a low copy number, whereas eight of them exhibited a high copy hybridization pattern. Four of these clones were converted into low copy probes by subcloning an intron segment of the corresponding gene by PCR. The position of the introns on the wheat ESTs was predicted aligning the coding sequence of wheat to the rice genomic DNA with the software SIM4 (http://gamay.univ-perp.fr/analyse_seq/sim4). Sequence tag site (STS) primers were designed on the EST exons flanking one of the introns as follows: STS 40 was designed on CD453029 (for. 5'-ggtgttttaagtgttttgcatgg-3'; rev. 5'-cttcaacatgactggaggacg-3'), STS 41 on BJ210832 (for. 5'-gaaagcttggaagaggcag-3'; rev. 5'-cccggttgtgccactgg-3'), STS 62 on BE516643 (for. 5'-gacatcgtgctgccggag-3'; rev. 5'-ccatccactcagagataatgacatcaa-3') and STS 77 primers are from BJ222628 (for. 5'-ggaccgcatcttcttcagcgt-3'; rev. 5'-atcttcttcttcttcttcatctatgcacc-3'). All the probes with a low copy hybridization pattern were first used to screen parental blots for polymorphisms, and were then hybridized to the *Ae. tauschii* BAC library. Parental blots included genomic DNA of the wheat cultivars 'Forno', 'Arina', 'Chinese Spring' and the Chinese Spring deletion line '7DS-4' (Endo and Gill 1996). DNA was digested with 27 endonucleases: *Alu*I, *Ava*II, *Dde*I, *Dpn*II, *Hae*III, *Msp*I, *Taq*I, *Bgl*I, *Nco*I, *Aci*I, *Dpn*I, *Hind*III, *Mse*I, *Sal*I, *Nde*I, *Kpn*I, *Cla*I, *Sac*II, *Bam*HI, *Dra*I, *Eco*RI, *Bgl*II, *Hind*III, *Xba*I, *Eco*RV, *Apa*I and *Sac*I. Digested DNA was

blotted on membranes and hybridized with ^{32}P-labeled probes to detect polymorphisms according to standard methods (Sambrook and Russell 2001).

BAC library screening, sequence analysis and genetic mapping

Screening of the *Ae. tauschii* BAC library was performed by hybridization as described by Wicker et al. (2003) using the low copy probes listed in table 1. Screening of the Glenlea BAC library was done by PCR according to Nilmalgoda et al. (2003) using the PCR markers STS 77, GWM1220 and GWM295. Subcloning of BAC DNA was done as described by Stein et al. (2000). Low pass sequencing of BAC clones was performed with ABI®3730 to identify genes and SSRs. Genes were detected with the BLASTn algorithm (Altschul et al. 1997). SSRs were identified using the SPUTNIK Perl Script (Abajian 2003). PCR primers were designed manually; primers were tested for annealing temperature and self complementarity with the software FastPCR (Kalendar 2006). Microsatellite markers were PCR amplified from wheat genomic DNA and visualized with the LiCor® DNA Sequencer 4200. The markers polymorphic between 'Arina' and 'Forno' were mapped in the 'Arina' x 'Forno' population, consisting of 240 single seed descent lines (Paillard et al. 2003). The linkage map was constructed using MAPMAKER 3.0b for MS-DOS (Lander et al. 1987). The new DNA markers on chromosome 7DS were integrated into the previously published 7DS genetic map (Schnurbusch et al. 2004b) using the "try" command and the new marker sequence was confirmed with the "ripple" command. Linkage groups were drawn with the software MapChart (Voorrips 2002). Physical mapping and bin assignment was conducted using the deletion line 7DS-4 (breakpoint 0.61) derived from cv. 'Chinese Spring', in which 39% of the short arm of chromosome 7D is missing (Endo and Gill 1996).

5.4 Results

Polymorphism analysis of wheat ESTs derived from the 'Lr34 *orthologous region' of rice and identification of Ae. tauschii BAC contigs*

The two EST markers BE493812 and BF473324 flanking the QTL *QLrP.sfr-7DS* had been previously located on rice chromosome 6S, where they define a physical contig of ~300 kb (Schnurbusch et al. 2004b). To increase marker density around the *Lr34* locus, we identified with a BLASTn search the wheat ESTs which are homologous to the open reading frames predicted in the rice orthologous region. The wheat ESTs were then hybridized to wheat genomic DNA to test for copy number and polymorphism, but no polymorphism was detected by any probe on the wheat D-genome in the cross 'Arina' x 'Forno'. To access potentially more polymorphic intergenic regions 19 low copy wheat ESTs and EST-derived probes (Table 1) were used to screen a BAC library of *Ae. tauschii* by hybridization.

Table 1. List of wheat probes detected with a BLAST search at the rice '*Lr34* orthologous region'. Only the probes with a low copy pattern were used to screen the *Ae. tauschii* BAC library.

Probe	Accession	Position in rice [kb]	Copy number	Number of identified *Ae. tauschii* contigs	Analyzed BACs from *Ae. tauschii*
WHE1275_C07_E13	BE493812	AP003708 [50-54]	low	1	
WHR21F21	BJ281290	AP003708 [160-164]	high	not used	
WHE1452_F03_K06 STS 40	CD453029	AP000399 [13-15]	high low for STS 40	1	
WH36C19 STS 41	BJ210832	AP000399 [16-19]	high low for STS 41	1	RI033N19

table 1 continues...

Probe	Accession	Position in rice [kb]	Copy number	Number of identified *Ae. tauschii* contigs	Analyzed BACs from *Ae. tauschii*
WHE1114_C04_E08	BE443044	AP000399 [21-23]	low	1	RI033N19
WHF22A08	BJ252160	AP000399 [48-53]	low	–	
WH19L04	BJ220373	AP000399 [88-91]	low	3	HD007J19
WHE0802_D01_G02	BE517741	AP000399 [93-98]	low	1	HD007J19
WHE618_B12_C24 STS 62	BE516643	AP000399 [106-109]	high low for STS 62	2	HD007J19
WHE4020_H05_P10	CA500527	AP000399 [134-137]	low	1	RI005C8[1] TaBAC1466J2
WHDL14M17 STS 77	BJ222628	AP003487 [14-20]	high low for STS 77	1	RI005C8[1] TaBAC1466J2
WHE0418_F09_K18	BE406581	AP003487 [42-50]	low	1	
WHE0923_B09_D17	BF473324	AP003487 [88-93]	low	1	BB045B13
CDO475	CDO475	AP003487 [79-86]	low	1	BB045B13
WHF1M10	BJ251589, 3x	AP003487 [56-57], [94-95], [98-99]	low	3	
FGAS053185	CV758803	AP003767 [13-14]	low	2	HD099L21
FGAS016080	CK163453	AP003767 [14-15]	low	1	HD099L21
BAGS32F04	BJ463527	AP003767 [36-38]	low	1	
HVU507094	AJ507094	AP003767 [44-49]	high	not used	
EBRO01_SQ003_A08	BI779107	AP003767 [58-64]	high	not used	

table 1 continues...

Probe	Accession	Position in rice [kb]	Copy number	Number of identified *Ae. tauschii* contigs	Analyzed BACs from *Ae. tauschii*
BAH11M12	AV835586	AP003767 [64-68]	low	1	HD099L21
TALR1130G03F	BG904265	AP003767 [68-70]	high	not used	
HA23F03	BU981349	AP003767 [95-99]	high	not used	
FGAS017807	CK206224	AP003632 [58-64]	low	1	

Table 1, footnote 1) RI005C8 is the biggest BAC clone of *Ae. tauschii* identified by hybridization with probes STS 77 and EST CA500527. STS 77 was also used as a PCR marker to screen the bread wheat BAC library of cv. 'Glenlea' and yielded TaBAC1466J2. Because these two BAC clones are derived from the same orthologous locus of chromosome 7 in *Ae. tauschii* and chromosome 7D in wheat cv. 'Glenlea', the BAC clone of 'Glenlea' was preferred to the one of *Ae. tauschii* for sequencing.

Seventy BAC clones were initially identified and 59 of them were found to be organized into eleven BAC contigs (Luo et al. 2003; http://wheatdb.ucdavis.edu:8080/wheatdb/). Five BAC contigs contained clusters of two to three ESTs. The corresponding genes in rice were also physically close to each other, with less than 10 kb separating them (Fig. 1). This suggests that these genes are collinear between rice and *Ae. tauschii,* although we cannot infer their relative physical distances in *Ae. tauschii* because a full BAC sequencing was not performed. One of the BAC contigs contained one cluster of genes as well as a single gene which are conserved between rice and *Ae. tauschii* and are separated by a region of ~300kb in *Ae. tauschii.* In rice they are separated by 110 kb. This BAC contig of *Ae. tauschii* is estimated to be ~1.2 Mbp (BAC contig containing the BAC clone HD099L21, Fig. 1).

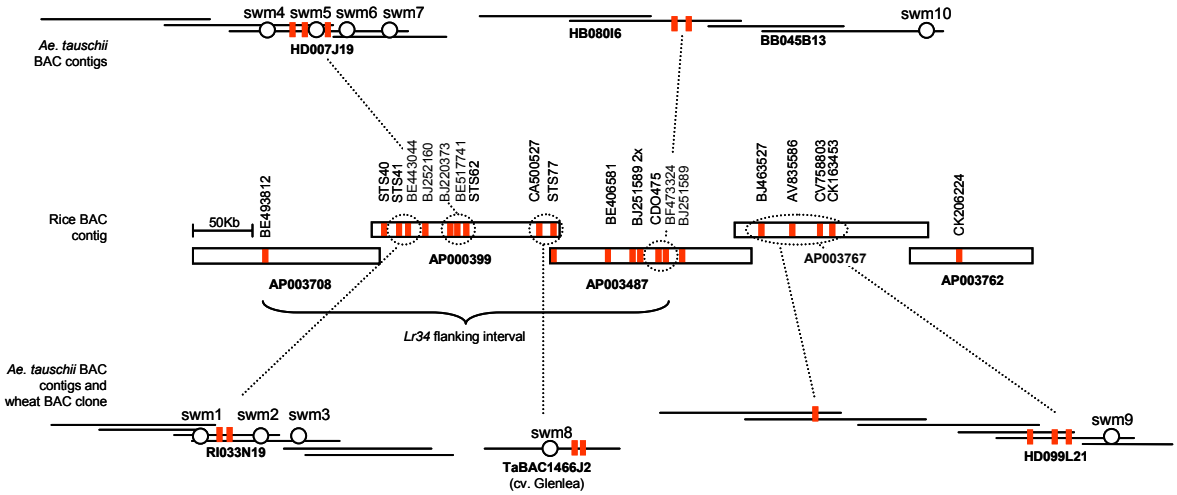

Figure 1. Identification of *Ae. tauschii* BAC contigs and wheat BAC clones from the targeted '*Lr34* orthologous region' of rice chromosome 6S. Only BACs from which new SSR were derived are shown. SSRs detected are indicated with a white circle. Low copy wheat, barley and oat EST and STS probes were used to screen the BAC library of *Ae. tauschii*. Probes are represented as grey boxes. Conserved gene clusters are framed by a dashed circle. The rice *'Lr34* orthologous region', delimited by ESTs BE493812 and BF473324, is marked with a curly bracket. TaBAC1466J2 from bread wheat cv. 'Glenlea' was preferred to the *Ae. tauschii* homolog to maximize SSR transferability. The relative sizes of the BAC contigs of *Ae. tauschii* are not drawn to scale.

For some low copy sequences we found gene redundancy in the *Ae. tauschii* genome. Five probes identified two or three BAC contigs, possibly derived from different genomic regions of *Ae. tauschii*. To increase the probability of selecting *Ae. tauschii* BAC clones from the chromosomal region of interest, we selected BAC clones independently identified by at least two probes. These were BAC clones BB045B13, HD099L21, RI033N19 and HD007J19 (Fig. 1).

Selection of BAC clones from chromosome 7D of the hexaploid wheat cv. 'Glenlea'

The BAC library of the hexaploid wheat cv. 'Glenlea' has a 3.5x genome coverage, an average insert size of 85 kb and is PCR screenable (Nilmalgoda et al. 2003). Since PCR markers are more specific than restriction fragment length polymorphism (RFLP) markers, they often allow tagging only one locus of the three orthologous genomes. Therefore, the microsatellite markers GWM1220, GWM295 and the sequence tagged site marker STS 77 were used to isolate BAC clones from the BAC library of 'Glenlea'. GWM1220 and GWM295 were described in previous studies to be linked to *Lr34* (Suenaga et al. 2003; Schnurbusch et al. 2004a, 2004b; Spielmeyer et al. 2005), and we could assign them to the bin 7DS-4 by deletion mapping (data not shown). STS 77 is an intron sequence of the EST BJ222628 (Table 1). This STS amplifies two bands of similar size, with the smaller assigned to bin 7DS-4. Only one BAC clone, TaBAC1466J2 with a size of ~85 kb, was found to contain the lower band of STS 77. Because TaBAC1466J2 was from bread wheat, it was preferred for sequencing to the corresponding BAC clone from *Ae. tauschii* (Table 1). Six BAC clones were identified with GWM1220 and three with GWM295. For these two markers, the BAC clones with the largest inserts were respectively TaBAC940L4 and TaBAC470M18.

Identification of microsatellite sequences in BAC clones of Ae. tauschii and bread wheat cv. 'Glenlea'

Three BAC clones of the 'Glenlea' D genome (TaBAC1466J2, TaBAC470M18 and TaBAC940L4) and four from *Ae. tauschii* (HD007J19, RI033N19, BB045B13 and HD099L21) were sub-cloned and low-pass sequenced to identify SSRs and coding sequences. From each BAC clone a shotgun library was derived and sequenced.

Four SSRs were detected in BAC clone HD007J19, three in BAC RI033N19, only one microsatellite was found in the BACs HD099L21, BB045B13 and TaBAC1466J2 (Table 2). In TaBAC470M18 and TaBAC940L4 no additional SSRs were detected besides GWM295 and GWM1220, which had been used for screening. A BLASTn analysis of the sequences flanking the simple sequence repeats revealed that these microsatellites were not located in known coding regions or repetitive elements. The size of the repeats ranged from 18 bp to 134 bp (Table 3). PCR primers were designed in the 5' and 3' regions flanking the repeat and were tested on genomic DNA of the lines 'Arina', 'Forno', 'Chinese Spring', and its deletion derivative '7DS-4'. SWM1, SWM5, SWM8, SWM9 and SWM10 produced bands specific for 7DS-4. SWM6 amplified two fragments, with one of them mapping to 7DS-4. SWM3, SWM4 and SWM7 were neither polymorphic, nor was it possible to physically assign them to a specific chromosome with the nullitetrasomic lines of Chinese Spring (Sears, 1966). Therefore, they were not considered for further analysis. Amplification of SWM2 was successful in the *Ae. tauschii* accession 'AL8/78', but it did not amplify a fragment in any wheat line (data not shown). This SSR was probably lost during the evolution of the wheat D-genome or mutations in the primer sequences occurred.

Table 2. Sequencing and molecular characterization of BAC clones from bread wheat cv. 'Glenlea' and *Ae. tauschii* identified with markers from the '*Lr34* orthologous region' of rice. The last column indicates the number of wheat predicted genes which were conserved in the orthologous rice region (see also Table 5).

BAC name – origin	Estimated BAC size	Total contigs length and number	SSRs	Contigs with repetitive DNA	Predicted genes	Predicted genes collinear in rice
RI033N19 – *Ae. tauschii*	150 kb	83 kb (56)	3	32	–	–
HD007J19 – *Ae. tauschii*	190 kb	160 kb (88)	4	55	4	1
TaBAC1466J2 – *T. aestivum*	85 kb	68 kb (25)	1	14	2	2
HD099L21 – *Ae. tauschii*	140 kb	84 kb (46)	1	21	7	4
BB045B13 – *Ae. tauschii*	140 kb	51 kb (85)	1	42	3	1
TaBAC940L4 – *T. aestivum*	120 kb	42 kb (64)	1 (GWM1220)	15	4	1
TaBAC470M18 – *T. aestivum*	110 kb	71 kb (35)	1 (GWM295)	16	3	3

Table 3. Ten microsatellite repeats were identified in the *Ae. tauschii* and wheat BAC clones of the rice '*Lr34* orthologous region'. Six of them amplified at least one fragment physically assigned to the bin 7DS-4 (in bold), where *Lr34* is located.

Probes used	BAC clones	*Lr34* SSR (7DS-4 deletion mapping)[1]	Motif	Forward primer (5'–3')	Reverse primer (5'–3')	Alleles detected in 17 lines
BE443044 STS 41	RI033N19	**SWM1 (y)**	$(tc)_{10}$	ACT CCC GAT ACA ATT CTT CCG CT	GAG ATC ATT GTA TTG AAG ATC AAA CG	6
BE443044 STS 41	RI033N19	SWM2 (n)	$(atag)_7$	ATG GAA TTC AAG AAT CGA TCT ACA C	GCG CTG TAA GCG AGA CAT AGG A	–
BE443044 STS 41	RI033N19	SWM3 (n)	$(cgc)_6$	TTC CTG AAC GAG CCT CTC C	CGG ACG ACT TGT TGT ATC ATA AGG	3

table 3 continues...

Probes used	BAC clones	*Lr34* SSR (7DS-4 deletion mapping)[1]	Motif	Forward primer (5'–3')	Reverse primer (5'–3')	Alleles detected in 17 lines
BJ220373 BE517741 STS 62	HD007J19	SWM4 (n)	$(ta)_{13}$	AGA TGC AGA TCT AAT GGT CAG AGA C	ACA ATA GCC CAC TTC TCT GCT C	1
BJ220373 BE517741 STS 62	HD007J19	**SWM5 (y)**	$(ta)_{32}/(tg)_{35}$	CCA GTA GCA TAA GCA TAA TAC AAC AC	ATA GAT TGT AAC ATT GCT TGA TTG C	9
BJ220373 BE517741 STS 62	HD007J19	**SWM6 (y)**	$(ac)_{11}$	CTT ACA CGG ACC ATG TGC AGA GG	AGG GAG TGG ATG AAC AAA GTG TG	7
BJ220373 BE517741 STS 62	HD007J19	SWM7 (n)	$(at)_{17}$	GTC GCA ATG TCA GAG AGG AAT CAG	GAG TGC ATT ACA CCA AGC TGT CTA G	1
STS 77	TaBAC1466J2	**SWM8 (y)**	$(ga)_{25}$	GCT CTT GAA CTT AGT CTC ATC AAG G	CTC TCC CGC TGC AGT GTC TC	3
CK163453 CV758803 AV835586	HD099L21	**SWM9 (y)**	$(gtt)_{14}$	GTC ATT CTG CCA ATG CAT GAT CC	GGT ACA GCC GAC ATA GGT CAT C	6
BF473324	BB045B13[2]	**SWM10[2] (y)**	$(ca)_{25}$	GCC TAC TTT GAC GGC ATA TGG	CCA TCT TGA CAT ACT TTG GCC TTC C	5

Table 3, footnote 1) y: present in 'Chinese Spring' but absent in the 7DS-4 deletion line;
n: not specific for the 7DS-4 deletion line.

Table 3, footnote 2) SWM10 was not identified from the BAC clone containing BF473324 but from a BAC clone of the same BAC contig.

Genetic mapping of microsatellites reveals a large-scale inversion between rice and wheat orthologous regions

All the newly developed microsatellites were tested for polymorphism between the parental lines for mapping in the 'Arina' x 'Forno' population (Paillard et al. 2003). SWM1, SWM5, SWM6 and SWM10 were polymorphic between the parents. SWM6 was assigned to the 7A linkage group. SWM1, SWM5 and SWM10 mapped in the 7D chromosome, closely linked to *Lr34*. However, in the wheat genetic map of chromosome 7D, the three SSR markers SWM1, SWM5 and SWM10 do not map inside the *Lr34* flanking interval, but outside and in a position more proximal to BF473324 (Fig. 2). Based on the assumption of wheat-rice collinearity, SWM1 should have been the marker mapping closest to EST BE493812, but in wheat it is the marker which maps most distant to it (18 cM). Similarly, SWM10 should have been the SSR with the largest distance to EST BE493812, but it is the closest (8.9 cM). SWM5 maintains its mapping position between SWM1 and SWM10, and it maps 13.6 cM proximal to BE493812. Thus, the genetic mapping in wheat suggests that the marker order in wheat is conserved, but in an inverted pattern compared to rice. This indicates the presence of a large inversion between rice and wheat '*Lr34* orthologous regions'. This hypothesis was further supported by sequencing the wheat BAC clone TaBAC470M18 which contains the SSR GWM295. This BAC clone was found to contain a gene encoding an early noduline-like sequence (CD904551). In the wheat genetic map, GWM295 maps between SWM5 and SWM1, within the inverted region. As expected according to the observed inversion, an orthologous early noduline-like gene was located in rice between orthologous sequences to the BAC clones RI033N9 and HD007J19. The genetic map was partially verified in a population of 600 'Arina' x 'Arina-*Lr34*' F2 plants (data not shown). Thus, due to the described inversion between wheat and rice, the new microsatellite markers mapped more proximal than expected assuming wheat-rice collinearity, and the microsatellite SWM10 is

the closest to the *Lr34* flanking interval that was defined by ESTs BE493812 and BF473324 by Schurbusch et al. (2004b).

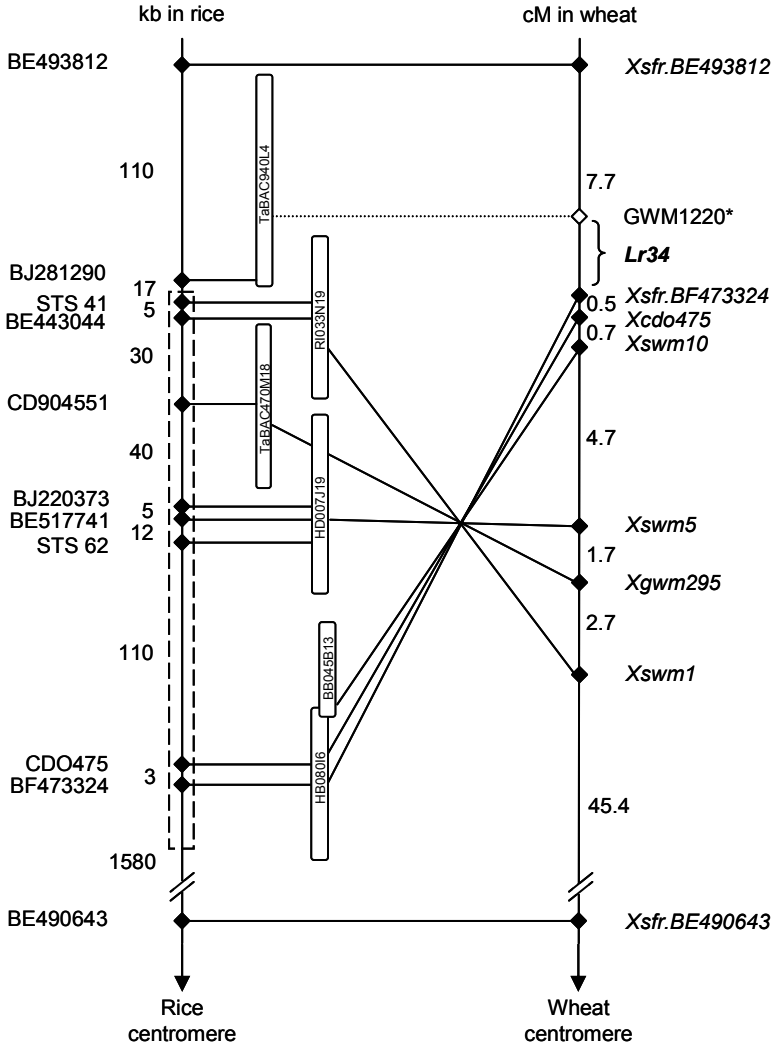

Figure 2. Genetic mapping of polymorphic microsatellites derived from BAC clones of the wheat cv. 'Glenlea' (TaBACs) and the wild grass *Ae. tauschii* allowed to correlate the genetic map of wheat chromosome 7D with the rice physical map of chromosome 6S. The inversion detected by mapping new SSR markers is indicated by a dashed box. BAC clones of wheat and *Ae. tauschii* are not drawn to scale. Only BAC clones HB086I16 and BB045B13 belong to the same contig, however they do not overlap. *Microsatellite GWM1220 was not polymorphic between 'Arina' and 'Forno'. Its mapping position is inferred from published work (Schnurbusch et al. 2004b; Spielmeyer et al. 2005).

Allele diversity study in wheat lines with and without Lr34

An allele diversity study was performed on a set of lines with and without *Lr34* using the newly developed SWM-SSR markers, as well as the two previously published SSR markers from the *Lr34* region GWM130 and GWM1220 (Suenaga et al. 2003; Schnurbusch et al. 2004b; Spielmeyer et al. 2005). The aim of this analysis was to check the degree of polymorphism at the new SSR loci in the wheat germplasm and to assess if any of the newly developed markers was diagnostic for the presence of *Lr34*. To assess the degree of polymorphism we used a set of lines extensively used to characterize *Lr34* (Table 4, 1st group). All microsatellites produced an amplification product in the lines except for SWM2, which amplified a fragment only in the genomic DNA of the *Ae. tauschii* accession 'AL8/78'. SWM4 and SWM7 were monomorphic, SWM3 and SWM8 produced three alleles; SWM1, SWM5, SWM6, SWM9 and SWM10 had a higher degree of polymorphism, ranging from three to nine alleles, comparable to GWM1220 and GWM130 (six and seven alleles respectively). It is noteworthy that the SWM5 allele in the introgression line'Arina-*Lr34*' is different from both the donor parent 'Forno' and the recurrent parent 'Arina' (data not shown). It seems that in a few generations of backcrossing (BC_2) and selfing (F_4) this *Lr34* line has developed a new allele at the *Xswm5* locus. Instability observed with SWM5 might depend on the size of this repeat. The part representing only the repeat of the amplified fragment of SWM5 is 134 bp long, and is very large compared to an average SSR. It consists of $(ta)_{32}/(tg)_{35}$ in *Ae.tauschii*, and the repeat might be very unstable, rapidly producing new allelic variants.

Table 4. Diversity analysis of bread wheat cultivars with the SSR marker SWM10. The allele 'a' is shared among the three independent sources of *Lr34*: 'Forno', 'Chinese Spring' and 'Frontana'. The genotypes in bold have *Lr34,* but do not have the *SWM10a* allele or vice versa, they have the *SWM10a* allele, but not *Lr34*. The alleles 'c', 'd', and 'e' appear with less frequency in the tested genotypes.

Line	*Lr34* (yes-no)	SWM10 allele	Line	*Lr34* (yes-no)	SWM10 allele
Model lines where *Lr34* was extensively characterized			**Elite lines with *Lr34***		
Arina	n	b	Tamaro	y	a
Forno	y	a	Tepoca	y	a
Arina *Lr34*	y	a	Terenzio	y	a
Chinese Spring (CS)	y	a	**Titlis**	**y**	**c**
CS 7D_Cheyenne	n	c	Tonichi 81	y	a
CS 7D_Hope	n	b	**Westphal 12A**	**y**	**d**
CS 7D_Red Egyptian	n	b	Yecora 70	y	a
CS 7D_Timstein	n	b	**Elite lines without *Lr34***		
CS 7D_Thatcher	n	b	Altgold	n	b
Lalbahadur	n	c	Apollo	n	b
Lalbahadur 7D_Parula	y	a	Ares	n	b
Jupateco 73S	n	d	Basalt	n	b
Jupateco 73R	y	a	Bobwhite	n	b
Avocet S	n	b	Boval	n	c
Avocet R	y	a	Can 3842	n	b
Thatcher *Lr16*	n	b	Champlein	n	b
Thatcher *Lr16 Lr34*	y	a	Derenburger	n	b
Elite lines with *Lr34*			Disponent	n	b
Bezostaja	y	a	Frisal	n	b
BH1146	y	a	Galaxie	n	d
Cappelle Desprez	**y**	**b**	Granada	n	c
Cumpas	y	a	Greif	n	b
Frontana	y	a	Hoeser	n	b
Glenlea	**y**	**e**	Hubel	n	b
Kormoran	**y**	**b**	Inia 66	n	d

table 4 continues...

Line	*Lr34* (yes-no)	SWM10 allele	Line	*Lr34* (yes-no)	SWM10 allele
Elite lines without *Lr34*			**Elite lines without *Lr34***		
Kanzler	n	b	Osmut	n	d
Kavkaz	**n**	**a**	Ostro	n	d
Kraka	n	b	Pavon 76	n	b
Kronjuwel	n	c	Pegassos	n	b
Little Club	n	b	Probus	n	b
Lueg	n	b	Rektor	n	b
Maris Huntsman	n	b	Roazon	n	b
Merlin	n	b	Rouquin	n	d
Monopol	n	c	Siete Cerros 66	n	b
Morocco	n	b	Skalavatis 56	n	b
Obelisk	n	b	Sperber	n	b
Oberkulmer	n	d	Weique	n	b

The microsatellite marker closest to the *Lr34* confidence interval is SWM10 (Fig. 2). Interestingly, in the set of *Lr34* model lines, the marker SWM10 showed an allele of the same size (allele 'a', 211bp) in all the lines with *Lr34,* but different alleles in the lines without *Lr34*. Therefore, this microsatellite was tested on a broader germplasm from different wheat breeding programs to determine its diagnostic value for detection of *Lr34*. In table 4, 2nd and 3rd section, additional elite genotypes are listed. Based on literature references it is also indicated whether they contain *Lr34* or not. The lines are mostly from the CIMMYT spring wheat germplasm and from the European winter wheat germplasm. To the best of our knowledge, none of the lines of the 2nd group has *Lr34*. Among this group, the only genotype having the *SWM10a* allele is 'Kavkaz'. Sawhney and Sharma (1999) reported that 'Kavkaz' manifests good adult plant leaf rust resistance. 'Kavkaz' is derived from a cross with the winter wheat line 'Bezostaja', which has been described to have *Lr34* (Winzeler et al. 2000),

therefore the line 'Kavkaz' might also contain *Lr34*, although this has not been published so far. Among the 3rd group of lines (Table 4), all described to contain *Lr34,* more variability in allele size was observed. A few lines that have been described to possess *Lr34* showed different bands than the 'a' SWM10 allele. These lines were the Canadian spring wheat cv. 'Glenlea' (reported to have *Lr34* by Dyck et al. in 1985) with the unique allele 'e' found in no other line, the German winter wheat line 'Westphal 12A' (Kolmer and Liu 2001), with allele 'd' and the French winter wheat line 'Cappelle Desprez' (McIntosh 1992), with allele 'b'. The German winter wheat line 'Kormoran' shows the 'b' allele as 'Cappelle Desprez' and also contains *Lr34* (McIntosh 1992). This could be explained by the presence of 'Cappelle Desprez' in the pedigree of 'Kormoran'. For the Swiss winter wheat cv. 'Titlis' (allele 'c'), there is no official report to support the presence of *Lr34;* nonetheless a strong adult leaf rust resistance and leaf tip necrosis lead to assume that this gene is present in 'Titlis' (Winzeler M., personal communication).

Putative coding regions and repetitive DNA in BAC clones of Ae. tauschii and wheat cv. 'Glenlea'

The analyzed BAC clones contained 23 putative genes, based on their homology with wheat ESTs from the TIGR database (Table 5). Nineteen sequences contained coding regions for proteins with predicted function. Two of them were homologous to the EST probes used to screen the *Ae. tauschii* BAC library and two others were similar to resistance genes of the NBS-LRR (nucleotide binding site – leucine rich repeat) class. In addition to these nineteen sequences with predicted function, four additional sequences had significant homology to wheat ESTs, but they did not show a hit to any characterized protein. Thus, we refer to them as 'hypothetical proteins'.

Table 5. A BLASTn analysis of the BAC clone sequences of wheat and *Ae. tauschii* reveals the presence of a number of sequences with homology to wheat ESTs and rice genes. Hits in the rice genome with conserved collinearity in the '*Lr34* orthologous region' are shown in bold. Hits not collinear are in plain text.

Homologous proteins	BAC	wheat ESTs	*E* value	Best rice hit
Cell wall associated kinase	HD007J19	BJ220373	$1.3xe^{-95}$	Os02g02120
Arginine N-methyl transferase[1]	**HD007J19**	**DN829072**	**$2.2xe^{-106}$**	**Os06g05090**
Hypothetical protein	HD007J19	AJ716967	$4.5xe^{-61}$	Os03g49250
NBS-LRR disease resistance homolog	HD007J19	CN011065	$7.4xe^{-88}$	Os08g28540
Sulfate transporter[1]	**TaBAC1466J2**	**CA685639**	**$6.4xe^{-32}$**	**Os06g05160**
Splicing factor PRP38	**TaBAC1466J2**	**CV765610**	**$2.2xe^{-141}$**	**Os06g05150**
Hypothetical protein	**HD099L21**	**CK212017**	**$5.5xe^{-79}$**	**Os06g05560**
GTP binding protein	HD099L21	BE213312	$4.8xe^{-15}$	Os09g19980
Proline-rich protein with lipase domain	**HD099L21**	**BE422772**	**$2.2xe^{-67}$**	**Os06g05550**
DNA helicase homolog	HD099L21	–	–	Os10g10730
Iron-sulfur cluster assembly accessory protein	**HD099L21**	**CA485358**	**$2.3xe^{-77}$**	**Os06g05400**
Mitogen-activated kinase	**HD099L21**	**CK214503**	**$3.1xe^{-70}$**	**Os06g05520**
Putative ABC transporter	HD099L21	–	–	Os03g64200
UDP-glucoronosyl and UDP-glucosyl transferase	BB045B13	CD871872	$2.4xe^{-19}$	Os01g50200
Zinc carboxypeptidase	**BB045B13**	**CK208877**	**$9.4xe^{-40}$**	**Os06g05240**
NBS-LRR disease resistance homolog	BB045B13	CA745220	$4.6xe^{-88}$	Os11g10770
Hypothetical protein	TaBAC940L4	CK207765	$1.9xe^{-19}$	Os08g06650
Hypothetical protein	TaBAC940L4	CK170616	$2.5xe^{-07}$	Os01g68130
O-methyltransferase	TaBAC940L4	CA633930	$4.9xe^{-64}$	Os11g20090
Serine-threonine kinase	**TaBAC940L4**	**BJ281290**	**$3xe^{-39}$**	**Os06g04880**
Early nodulin	**TaBAC470M18**	**CD870012**	**$1.3xe^{-105}$**	**Os06g05010**
Early nodulin	**TaBAC470M18**	**BQ167397**	**$2.3xe^{-73}$**	**Os06g05020**
Oxidoreductase, short chain	TaBAC470M18	CV761584	$1.3xe^{-59}$	Os06g19590

Table 5, footnote 1) EST probes with homology to these predicted genes were used to screen the BAC library of *Ae. tauschii*.

The 23 predicted coding sequences were analyzed with BLASTn to identify the homologous rice genes and their localization in the rice genome. Collinearity with rice was conserved only for 11 of them, while 12 sequences, including some genes putatively encoding hypothetical proteins or resistance gene analogues of the NBS-LRR class, were not conserved in the rice orthologous region. Earlier studies have shown that in cereals, disease resistance genes of the NBS-LRR family are subject to rapid evolution and often do not maintain collinearity (Leister et al. 1998).

Physical to genetic distances at the Lr34 *locus*

The three molecular markers BF473324, CDO475 and SWM10 were polymorphic on chromosome 7D in the 'Arina' x 'Forno' population, and were derived from the same *Ae. tauschii* physical contig, which spans ~700 kb (Fig. 3). This allowed anchoring and orientation of a large DNA sequence of *Ae. tauschii* BAC clones on the wheat genetic map in chromosome 7D. In this region, the relationship between physical and genetic distances was assessed based on BAC hybridization information. The ESTs BF473324 and CDO475 were located in a ~20 kb region (Fig. 3). SWM10 was separated from them by ~560 kb of DNA. Thus, overall physical to genetic relationship is ~500 kb/cM. However, this ratio varied in the region, since the 0.7 cM genetic interval defined by *Xswm10* and *Xcdo475* corresponded to 560kb, while the adjacent genetic interval of 0.5 cM, defined by *Xcdo475* and *Xsfr.BF473324* was only ~20 kb (Fig. 3). These two ESTs correspond to genes that in wheat have maintained a similar distance as in rice. The differences in the kb/cM ratio could possibly be explained by variations in gene content of the two intervals.

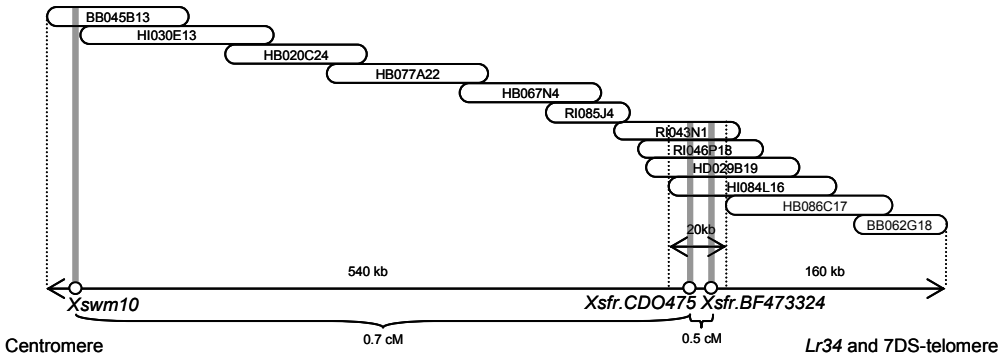

Figure 3. Genetic to physical distances in the *Lr34* region. On the upper side a selection of BAC clones from the *Ae. tauschii* BAC contig4981 is represented, according to the assembly of March 7th 2004 (http://wheatdb.ucdavis.edu:8080/wheatdb/). Below, white circles show the mapping position of three molecular markers developed from this BAC contig in the wheat 7DS chromosome. This allowed orienting the *Ae. tauschii* BAC contig on the 7D chromosome, and establishing a relationship between physical and genetic distances at the *Lr34* region.

5.5 Discussion

Tagging of a specific region in the D-genome of hexaploid wheat

In this work we have tested for polymorphism a set of 24 ESTs markers derived from the rice '*Lr34* orthologous region' to saturate the *Lr34* containing interval in wheat. No RFLP was detected in the D-genome between the parental lines 'Arina' and 'Forno'. This outcome is consistent with the low degree of polymorphism in the D-genome of cultivated bread wheat. SSR markers represent a good alternative for mapping because of their high level of polymorphism (Prasad et al. 2000; Singh et al. 2006). The identification of SSR markers in *Ae. tauschii* BAC clones allowed the enrichment of the *Lr34* region on 7DS with new polymorphic microsatellite markers. Eight of the nine microsatellites derived from the *Ae. tauschii* genome could be transferred to the D-genome of bread wheat (SWM8 was derived from the D-genome of 'Glenlea', SWM2 was not transferable). The *Ae. tauschii* BAC clones have been successfully used as a bridge to link the physical information of the rice sequence to genetic mapping in wheat. This strategy is of general utility for the development of markers for gene tagging in the D-genome of hexaploid wheat.

Low pass sequencing of BAC clones revealed a high density of SSR motifs. In a total of 560 kb of sequence, ten SSRs were identified, giving an average of one microsatellite every 56 kb. Pestsova et al. (2000) reported microsatellite isolation from *Ae. tauschii* phage libraries by hybridization with the dinucleotide repeats poly(GA) and poly(GT) and found a repeat every 220 kb. In our case sequence data were available, and this allowed the detection of all the possible repeat combinations with an *in silico* approach. This explains the four fold higher density of SSRs found.

Most of the developed SSRs tagged specifically the 7DS-4 bin. This was surprising as we could not exclude an occasional amplification from paralogous loci in the D-genome. In a few cases, the ESTs derived from the rice *Lr34* orthologous region hybridized with BAC clones assigned by BAC fingerprinting to different BAC contigs, suggesting that they are located in paralogous regions of the D-genome. To increase the chances to tag the BACs belonging to the *Lr34* region, we selected only BAC clones which were hybridizing with probes of two or more genes closely associated in rice (Table 1). Obviously, this strategy successfully avoided the erroneous targeting of paralogous loci in the D-genome. Using BACs of diploid *Ae. tauschii* we also reduced the chances of targeting orthologous loci in the A and B genomes of bread wheat, as described by Shen et al. (2005), while developing SSR markers on a genome wide scale.

Potential and limitations of the rice genome for high resolution analysis of the wheat genome

Rice is the only grass species with a completely known genome sequence. Comparative genetic studies revealed that within the grass family there is good collinearity and, despite major rearrangements of big chromosomal segments, the order of conserved genes inside each block is frequently maintained (Bennetzen and Ma 2003). Wheat chromosome 7D is partially collinear to the rice chromosome 6. By mapping SSR and EST markers derived from the rice '*Lr34* orthologous region', we observed a large genomic inversion including part of the *Lr34* containing interval. Small inversions affecting microcollinearity have been described earlier in Triticeae genomes (Dubcovsky et al. 2001; Isidore et al. 2005). Furthermore, in the *Ae. tauschii* and wheat BAC clone sequences, we identified a number of sequences with homology to ESTs that were not collinear to the targeted region of rice. These results are in

agreement with earlier findings of frequent distortion of microcollinearity between rice and Triticeae genomes (Bennetzen and Ramakrishna 2002; Feuillet and Keller 2002; Akhunov et al. 2003). This mosaic conservation of microcollinearity between rice and Triticeae genomes is complicating the use of the rice model genome. However, as described below, the information from the rice genome can still provide very relevant information on candidate genes. For targeted SSR development, the application of rice genome information for the isolation of BAC clones/contigs of *Ae. tauschii* was extremely useful to saturate with markers a specific region of the wheat genome. A more complete physical map of wheat and/or *Ae. tauschii* and a better whole-genome knowledge at high resolution of rice/Triticeae genomic rearrangements would provide the necessary, improved tools for efficient work on the wheat genome.

Definition of a candidate region for Lr34

The allelic study at the *Xswm10* locus indicates that this marker is genetically close to the *Lr34* region. Microsatellite mapping revealed a large genomic inversion involving part of the *Lr34* flanking interval. According to this information, the rice genomic region syntenic to *Lr34* is now supposed to be proximal to EST marker BF473324. In this rice region, two related genes, both annotated as pectate lyase genes, were detected. Three wheat ESTs covering different parts of these rice genes were identified (data not shown). In *Arabidopsis,* the powdery mildew resistance gene *PMR6* has a pectate lyase structure and shows good similarity at the protein level with an assembly of the three wheat ESTs. At the protein level, a 73% similarity was found in a sequence of 268 amino acids. In *Arabidopsis,* a mutation in the gene *PMR6* was described to confer horizontal resistance to powdery mildew (Vogel et al. 2002). Analysis of mutants revealed that *PMR6* in *Arabidopsis* alters cell wall composition

and has a pleiotropic effect on leaf morphology, providing a pre-haustorial kind of resistance (Vogel et al. 2002). The *Lr34* locus also confers a pre-invasion resistance (Niks and Rubiales 2002), is effective against powdery mildew (Spielmeyer et al. 2005) and has pleiotropic effects on leaf morphology, e.g. it is cosegragating with leaf tip necrosis. Pectate lyases might play a central role in remodeling the structure of the cell wall. Therefore, they might exert a cell wall based resistance by delaying the haustorial contact with the symplast that biotrophic plant pathogen need to establish. For this reason, we speculate that the wheat pectate lyase locus in 7DS could possibly be involved in the horizontal resistance conferred by *Lr34*.

Three independent sources of Lr34 *'Frontana', 'Chinese Spring' and 'Forno' share a common allele at the* Xswm10 *locus*

The *SWM10a* allele is common to the three known independent sources of *Lr34.* This allele, which has a fragment size of 211 bp, is identical in the genotypes 'Frontana', 'Chinese Spring' and 'Forno'. To our knowledge, these cultivars do not share pedigree, although genetic distances of the tested genotypes are not known. The *SWM10a* allele is also common to all tested spring wheat lines developed at the CIMMYT, in which *Lr34* was derived from 'Frontana'. Nonetheless, in the European winter germplasm with *Lr34*, more variability at the *Xswm10* locus was observed. Genetic recombination, partial SSR instability or seed impurity could be involved in the observed discrepancies between presence of *Lr34* and absence of the *SWM10a* allele in some of these genotypes. Thus, although SWM10 seems not to be completely diagnostic when using a wide germplasm, it will be an extremely useful marker to assist breeding programs for *Lr34/Yr18* worldwide. Finally, the knowledge on the orientation of the SWM10 BAC contig on the 7D chromosome by genetic mapping provides the basis for chromosome walking towards *Lr34*.

Acknowledgements:

We would like to thank Dr. Thomas Wicker and Dr. Edith Schlagenhauf for their bioinformatical support and to Beatrice Senger, Theres Klimm and Urs Ochsner for their technical assistance. We are very thankful to Dr. Jan Dvorak and Dr. Alina Akhunova (University of California, Davis, CA, USA) for providing BAC clones and filters of the *Ae. tauschii* BAC library, and to Dr. Sylvie Cloutier and Dr. Andrzej Walichnowski (Cereal Research Centre, Winnipeg, Canada) for providing the BAC pools of the 'Glenlea' library as well as wheat ESTs. For providing ESTs we also thank Dr. Olin Anderson and Dr. Curt Crossman (USDA-ARS, Albany, CA, USA), Dr. Yasunari Ogihara (Yokohama City University, Japan), Dr. Kazuhiro Sato (Okayama University, Japan), Dr. Mark Sorrells (Cornell University, Ithaca, NY, USA), Dr. Jens Stougaard (Aarhus University, Denmark), Dr. Patrick Gulick (Concordia University, Montreal, Canada), Dr. Nils Stein (IPK Gatersleben, Germany), Dr. Robert Waugh (SCRI, Dundee, Scotland, UK). We thank Dr. Marion Röder (IPK Gatersleben, Germany) as well as TraitGenetics Inc. (Gatersleben, Germany) for providing unpublished primer sequences. We are grateful to Dr. Jon Raupp (KSU, WGRC, Manhattan, KS, USA) for providing plant material. Finally we would like to thank Dr. Nabila Yahiaoui and Dr. Jan Dvorak for critical reading of the manuscript. This work was supported by the Swiss National Science Foundation (SNF grant 3100A0-105620) and by the Indo-Swiss Collaboration in Biotechnology (ISCB, project WN7). The authors declare that the experiments conducted for this publication comply with the current laws of Switzerland.

Chapter 6

COMPARISON OF ORTHOLOGOUS LOCI FROM SMALL GRASS GENOMES *BRACHYPODIUM* AND RICE: IMPLICATIONS FOR WHEAT GENOMICS AND GRASS GENOME ANNOTATION

* Eligio Bossolini, * Thomas Wicker, Philip Knobel and Beat Keller

* *both authors contributed equally to the work*

(2006)

the Plant Journal, submitted for revision

6. COMPARISON OF ORTHOLOGOUS LOCI FROM SMALL GRASS GENOMES *BRACHYPODIUM* AND RICE: IMPLICATIONS FOR WHEAT GENOMICS AND GRASS GENOME ANNOTATION

6.1 Abstract

Brachypodium sylvaticum and *distachyon* were recently proposed as new model plants because of their small genomes and their phylogenetic position between rice and Triticeae crops. We sequenced a 371 kb region in *B. sylvaticum*, the largest genomic sequence available from this species so far, providing quantitative data on gene conservation, colinearity and phylogeny. We compared it with orthologous regions from rice and wheat. *Brachypodium* and wheat show perfect macro-colinearity of genetic markers, whereas rice contains a ~220 kb inversion. Rice contains almost twice as many genes as *Brachypodium* in the region studied whereas wheat has about 40% more. Through comparative annotation we identified alternative transcripts and improved the annotation for several rice genes, indicating that ~15% of rice genes might require re-annotation. Surprisingly, our data suggest that 10-15% of functional sequences in small grass genomes may not encode any proteins. From available genomic and EST sequences we estimated *Brachypodium* to have diverged from wheat about 35-40 MYA, significantly less than the divergence of rice and wheat. However, our data also indicate that orthologous regions from *Brachypodium* and wheat differ considerably in gene content and, thus, the *Brachypodium* genome sequence can probably not replace genomic studies in the large Triticeae genomes.

6.2 Introduction

Brachypodium sylvaticum and *distachyon* are wild grasses that were recently proposed as a new model organisms for structural and functional genomics of temperate cereals because of their small genomes of approximately 400 million bp (Mbp) respectively and short reproduction cycle (Draper *et al.*, 2001; Foote *et al.*, 2004; Hasterok *et al.*, 2006). Efforts to completely sequence the *Brachypodium distachyon* genome are underway and a first draft genome sequence is expected in the near future. Phylogenetic studies placed the genus *Brachypodium* closer to the cool season crops than to rice (Catalan and Olmstead, 2000). Thus, the *Brachypodium* species were suggested to bridge the "genomic gap" between rice and the Triticeae family which contains some of the worlds most important crops such as wheat and barley (Draper *et al.*, 2001, Foote *et al.*, 2004).

It is estimated that rice and the Triticeae/*Brachypodium* lineage diverged approximately 50 million years ago (MYA, Paterson *et al.*, 2004). Wheat and barley diverged approximately 10-14 MYA whereas the different wheat species evolved within the last 2-7 million years (Wolfe *et al.*, 1989). However, it is still debated where precisely *Brachypodium* has to be placed in the phylogenetic tree and how many million years ago this genus diverged from the Triticeae lineage. The Triticeae tribe includes over 200 species. They have large diploid genomes of >5,000 Mbp which contain at least 80% repetitive DNA (Bennett and Smith, 1976), distinguishing them clearly from *Brachypodium* and rice with their much smaller genomes. Several Triticeae species are polyploids (e.g. hexaploid wheat with A, B and D-genomes) which adds to the size and complexity of their genomes. Despite their great economic importance, only a limited set of genomic sequences is available. As of May 2006, a total of 174 genomic sequences longer than 10 kb were deposited at GenBank. The total cumulative

size of these sequences (16.3 Mbp) represents less than 0.3% of a haploid genome equivalent. The lack of a physical map of a Triticeae genome requires to isolate genetically defined genes of interest by time-consuming map based strategies.

It has been shown with molecular markers that grass genomes display a high degree of genetic colinearity, reflecting their descent from a common ancestor (Gale and Devos, 1998). The overall conservation of gene content and linear order of genes is described as macro-colinearity. At this level of analysis, strong conservation of large chromosome segments of hundreds or thousands of kb is apparent, also because markers mapping in non-orthologous locations are usually considered to be paralogous loci (Bennetzen, 2000). Because of its compact genome and agronomic importance, the rice genome was the first grass genome to be completely sequenced (International Rice Genome Sequencing Project, 2005), and it was proposed as a model genome to study the larger and more complex grass genomes by serving as an "anchor" genome. Strong conservation of gene order between rice and wheat across large chromosomal regions was demonstrated recently at the *Ph1* locus in wheat (Griffiths *et al.*, 2006). However, numerous studies also show that colinearity frequently breaks down when sequences are compared at the level of individual genes. At this level of micro-colinearity, a mosaic conservation of genes between grass species is observed with some genes being present in both and some in only one species (Bennetzen and Ma, 2003; Dubcovsky *et al.*, 2001; Guyot and Keller, 2004; Song *et al.*, 2002; Swigonova *et al.*, 2005; Tarchini *et al.*, 2000). Translocations, deletions and duplications of genes cause these frequent exceptions in colinearity (Bennetzen and Ma, 2003; Feuillet and Keller, 2002). Especially genes of economic importance such as resistance genes are usually not found in colinear positions, limiting the usefulness of rice as an anchor genome for gene cloning (Gallego *et al.*, 1998; Leister *et al.*, 1998).

The precise extent of micro-colinearity between grasses is unknown because only the rice genome sequence is available in its entirety and because many annotated genes might actually be transposable element (TEs) sequences (Bennetzen *et al.*, 2004). Recent studies have shown that TEs such as Pack-MULEs, helitrons and CACTA transposons can carry fragments of genes and, thus, produce the impression of frequent interruptions in colinearity (Jiang *et al.*, 2004; Morgante *et al.*, 2005; Wicker *et al.*, 2003a).

The many exceptions found in micro-colinearity between rice and the genomes of wheat and barley led to an increased interest in the genome of *Brachypodium.* Molecular tools such as BAC libraries were developed from both *Brachypodium sylvaticum* (Foote *et al.* 2004) and *B. distachyon* (Gu *et al.* 2006; Hasterok *et al.* 2006) and a large resource of more than 20,000 ESTs from *B. distachyon* became available recently (Vogel *et al.*, 2006). However, genomic sequence resources are extremely limited: as of June 2006, the (unfinished) sequences of only three BAC clones were publicly available. The recent characterization of the *Ph1* locus demonstrated by hybridization analysis that *B, sylvaticum* genes are more similar to wheat than to rice, implied that colinearity between wheat and *Brachypodium* is better than between wheat and rice and showed that *Brachypodium* BAC clones can be used to generate new genetic markers (Griffiths *et al.*, 2006). However, more molecular data and extensive comparative analysis are needed to evaluate genome micro-colinearity of *Brachypodium* with Triticeae genomes.

Here, we present the analysis of 371 kb genomic sequence from *B. sylvaticum* and its comparison with orthologous sequences from rice and wheat. We found that gene order and content are conserved to a considerable degree between the three species but numerous exceptions were also found. There are several characteristics that place *Brachypodium* clearly closer to the Triticeae than to rice, such as the presence of a large inversion in rice as well as

the conservation of gene paralogs specific to the *Brachypodium*/wheat lineage. However, our data also indicate that *Brachypodium* and Triticeae have diverged approximately 35-40 MYA and that micro-colinearity between their genomes is quantitatively similar to their micro-colinearity with rice.

6.3 Materials and Methods

Shotgun sequencing

BAC DNA was isolated with the QIAGEN large construct kit. After mechanical shearing and size selection on an agarose gel, 3-5 kb fragments were ligated into the Topo-Blunt vector (Invitrogen). Electrocompetent DH-10B *E. coli* were used for transformation. Plasmid DNA was isolated in 96-well plates on a QIAROBOT (QIAGEN) and sequenced on an ABI3730 automated sequencer (Applied Biosystems) by the Sanger method. Base calling and quality trimming of the sequences was done using PHRED (Ewing *et al.*, 1998) and the initial assembly of BAC sequences was done with the PHRAP assembly engine (version 0.990319, provided by P. Green and available at www.phrap.org). Gaps in the BAC sequence were closed by PCR amplification directly on the BAC clones.

Sequence analysis

For sequence analysis BLAST (Altschul *et al.*, 1997), CLUSTALW (Thompson *et al.*, 1994) and DOTTER (Sonnhammer and Durbin, 1995) were used. Stand alone Blast software was

obtained from NCBI (www.ncbi.nih.gov). For analysis of rice sequences, datasets from TIGR rice genome (version 4, www.tigr.org) were used. All Triticeae, rice and *Brachypodium* ESTs were downloaded from NCBI. For the annotation of *Brachypodium* genes, the rice CDS with the highest identity was aligned with the *Brachypodium* genomic sequence using DOTTER to determine the positions of introns, exons, start and stop codons. For a comparison between Triticeae, rice and *Brachypodium* CDS Triticeae ESTs were identified by BLASTN search against all Triticeae ESTs. Triticeae ESTs were aligned with the *Brachypodium* CDS. If ESTs covered only parts of the *Brachypodium* gene, ESTs were assembled manually into contigs to cover as much of the gene as possible. A consensus sequence was made if several ESTs were available which covered the whole gene. The corresponding rice, *Brachypodium* and Triticeae gene sequences were aligned using CLUSTALW (Thompson *et al.*, 1994). The resulting multiple sequence alignment was used for pairwise comparison and calculation of sequence identity. Positions with gaps were ignored. Repetitive elements identified on the *Brachypodium* sequence were submitted to the TREP database (http://wheat.pw.usda.gov/ITMI/Repeats). Phylogenetic analysis was performed with the PHYLIP package (http://evolution.genetics.washington.edu/phylip/) using the protein sequence parsimony method (PROTPARS) on 1000 bootstrap replicates with jumbling the order of sequences 3 times for each replicate. The alignment used for the analysis is provided (Supplementary Figure 2).

Linkage mapping

Mapping on wheat chromosomes 7A and 7D was done using the 'Arina' x 'Forno' recombinant inbred lines (Paillard *et al.*, 2003). The SWM microsatellite markers were developed and mapped as described in Bossolini *et al.* (in press). Wheat ESTs and the probe

SGC77 were mapped as RFLP probes using the hybridization conditions described by Stein *et al.* (2000). The linkage map was constructed using MAPMAKER 3.0b for MS-DOS (Lander *et al.*, 1987). The new DNA markers on chromosome 7D and 7A were integrated into the previously published genetic map (Schnurbusch *et al.* 2004, Paillard et al 2003) using the TRY command and the new marker sequence was confirmed with the RIPPLE command.

6.4 Results

Sequencing of BAC clones from the Lr34 orthologous region of Brachypodium and sequence annotation

We have recently described the genetic mapping of the *Lr34* locus in wheat (Schnurbusch *et al.*, 2004) and the identification of its orthologous region on rice chromosome 6 (Bossolini *et al.*, in press). For several rice genes from this region, wheat ESTs were identified which where then used to design probes for genetic fine mapping in wheat. Here, we used three of these wheat ESTs (*BE483812*, *BF473324* and *BE443044*) to screen the BAC library of *B. sylvaticum* (Foote *et al.*, 2004). The 10 BAC clones identified by the probes belong to a single contig as shown by specific PCR with primers derived from BAC end sequences (Figure 1). Three BAC clones (58B21, 28A13 and 34I17) were selected for complete sequencing. 28A13 and 34I17 overlap by little more than 6 kb whereas 58B21 and 28A13 do not overlap at all. However, restriction mapping of the ten BAC clones showed that the gap between the two must be small and it could be closed by PCR. Indeed, the gap turned out to be only 1.4 kb. The resulting combined sequence has a length of 371'567 bp and is referred to as *Bsil_Lr34*.

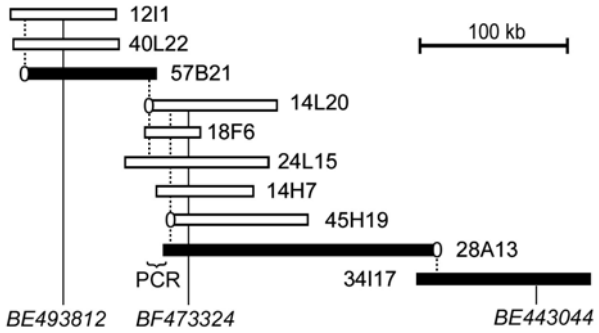

Figure 1. Map of the *Brachypodium* BAC contig spanning the segment orthologous to the wheat *Lr34* region. Locations of ESTs used for the screening of the BAC library are indicated with vertical lines. BAC ends used to identify overlaps are marked with ovals and BACs which contain these ends are connected with dotted lines. A 1.4 kb gap between the non-overlapping BAC clones 57B21 and 28A13 was closed with PCR on BAC and genomic DNA (indicated with a curly bracket). BACs chosen for complete sequencing are depicted as solid bars.

The precise borders of the region in the rice genome orthologous to *Bsil_Lr34* was identified by BLASTN of the *Bsil_Lr34* sequence against the TIGR rice genome (version 4). The region orthologous to *Bsil_Lr34* is located on the subtelomeric segment of rice chromosome 6S, approximately at positions 2,026 kb – 2,407 kb and is, thus, very similar in size (approximately 380 kb). For comparative analysis a fragment ("*Osat_Lr34*") from chromosome 6 positions 2,010,000 – 2,410,000 was used. All bp positions from rice described in the following refer to that fragment. *Osat_Lr34* contains 72 annotated genes according to the TIGR annotation.

DotPlot alignment of the rice and *Brachypodium* sequences showed a high degree of conservation in the linear order of sequences between the two species (represented as scattered diagonal lines in the DotPlot alignment, Figure 2a). The most striking difference in macro-colinearity between the two sequences is a large inversion of more than 200 kb (Figure 2a). However, several interruptions in micro-colinearity (interruptions in the diagonal lines) are also visible.

The first step of the annotation of *Bsil_Lr34* was the identification of gene candidates based mainly on rice genes predicted for the *Osat_Lr34* region. In this phase, we also annotated apparent pseudogenes and gene fragments because they still can provide information on colinearity and evolutionary relationships. The sequences conserved between the two species largely correspond to coding sequences (CDS) of genes whereas introns as well as up- and downstream sequences have diverged to a degree that virtually no sequence similarity can be detected anymore. For example, DotPlot comparison of a gene encoding a coatomer beta subunit (*Os06g05180*) with its ortholog from *Brachypodium* shows that all 23 exons of the gene are strongly conserved and can easily be identified (Supplementary Figure 1a). Thus, for the annotation of most genes, the predicted CDS of rice genes could be used to predict the location and exon/intron structure of *Brachypodium* genes. In *Brachypodium* regions which have no counterpart in the *Osat_Lr34* sequence, we identified by BLASTN and BLASTX eight candidate genes that are non-colinear and have their homologs elsewhere in the rice genome.

Some candidate genes are conserved non-coding sequences

Five regions are conserved between *Bsil_Lr34* and *Osat_Lr34* at colinear positions which appear to be non-coding conserved sequences (CNS) since they show no homology at the protein level to known genes and no open reading frames could be identified. CNS are considered potentially functional sequences and part of the "gene space" of a genome (Kaplinsky *et al.*, 2002). Thus, we included them in the annotation as candidate genes, bringing the total number of candidate genes to 47 for the *Brachypodium Bsil_Lr34* and 77 for rice *Osat_Lr34* sequences, respectively. Two of the CNS (*TRA-1* and *TRG-1*) are genes for $tRNA^{Ala}$ and $tRNA^{Gly}$, respectively. The other three (*CR-1*, *CR-2* and *CR-3*) are of unknown

function. For none of them, ESTs from rice, *Brachypodium* or wheat could be identified. DotPlot alignment shows that *CR-1* forms of a string of several conserved motifs that extends over almost 5 kb, indicating the exon/intron structure of a putative gene (Figure 2b). It is possible that this region contains a novel type of gene (protein-coding or not) that is conserved in grasses and which has escaped identification due to the absence of ESTs and due to its lack of similarity to known genes. *CR-2* and *CR-3* are 152 bp and 136 bp, respectively, and ~80% identical between the two species (Supplementary Figure 1b and 1c). The fact that both sequences are single copy in the rice genome and are found at colinear positions in the two species suggests them to be functional motifs.

Additionally, DotPlot showed the presence of short 15-30 bp motifs in the upstream and downstream sequences of most genes that are conserved between the two species. Most of these motifs were found at similar distances from the predicted CDS in both species and may be conserved regulatory elements in the otherwise highly divergent promoter or downstream regions of conserved genes (data not shown).

Regions that remained un-annotated at this point in the *Brachypodium* sequence were searched for the presence of repetitive DNA elements. We identified a tota l of 40 MITEs, seven LTR- and six non-LTR retrotransposons as well as ten Mutator transposons. Retrotransposons and Mutator elements were identified by homology of CDS to known repeats as well as through the identification of structural features such as long terminal repeats (LTRs) or terminal inverted repeats (TIRs). Six repetitive elements could not yet be classified. In total, 29% of the *Bsil_Lr34* sequence were classified as repetitive DNA. It can be expected that with increasing sequence data from *Brachypodium*, more repetitive elements will be discovered and that the actual repeat content is higher. Additionally, some of the annotated candidate genes possibly belong to TEs (see below).

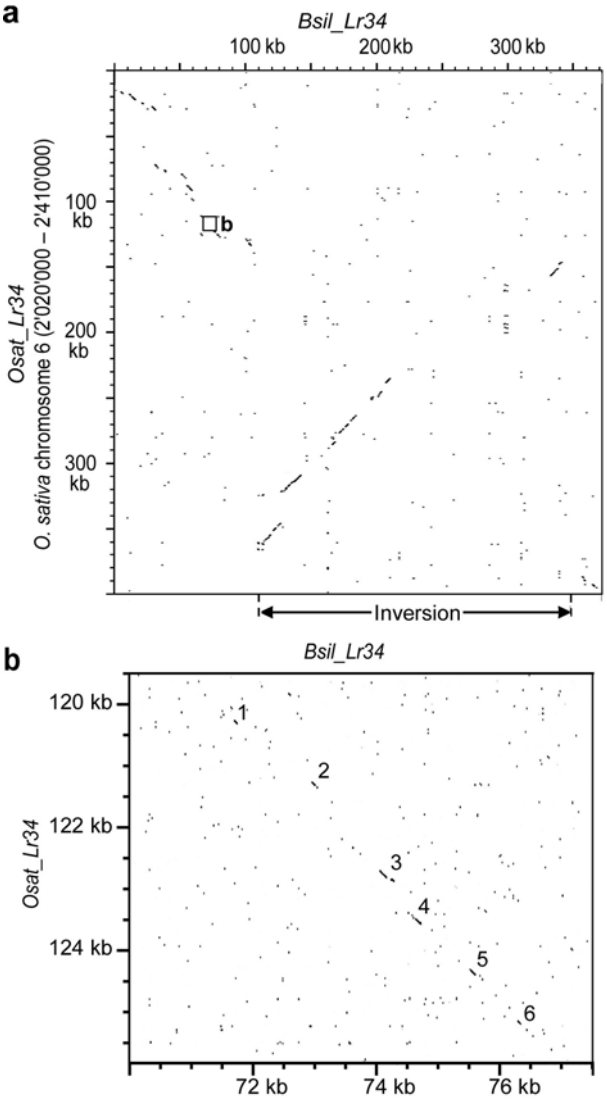

Figure 2. DotPlot alignment of the *Lr34* orthologous region from *Brachypodium* (horizontal) and rice (vertical). **a.** The overall conservation of coding sequences is represented as scattered diagonal lines. The locus contains a large inversion of more than 200 kb. The square indicates the location of the region depicted in b. **b.** DotPlot alignment of *CR-1* a conserved putatively non-coding sequence (CNS). No genes could be predicted or ORFs identified in this region. *CR-1* consists of a series of conserved motifs, suggesting either six independent CNS or the presence of a gene with at least 6 exons.

Several candidate genes are parts of transposable elements

Previous studies showed that coding regions of TEs in grass genomes are often mistaken for genes (Bennetzen *et al.*, 2004; Morgante *et al.*, 2005; Wicker *et al.*, 2003a). Thus, we defined

criteria to distinguish "real" genes from those that are likely to be part of TEs in order to obtain a conservative estimate of the actual gene content of both sequences. All sequences that are conserved in colinear positions between *Brachypodium* and rice were considered to be real genes whereas all non-colinear candidate genes were examined using the additional three criteria discussed below:

- First, if a candidate gene either encodes TE proteins or was found within typical TE boundaries such as LTRs or TIRs it was classified as part of a TE. Six genes from rice met this criterion and none from *Brachypodium*, since we identified several typical TEs already in earlier steps of the *Brachypodium* annotation (see above).
- Second, many TEs evolve rapidly and are hardly conserved between distantly related species such as rice and Arabidopsis. However, many TEs are still conserved (at least at the protein level) between grasses due to closer phylogenetic relationships and are often annotated as genes.Thus, we used all non-colinear candidate genes as queries in a BLASTX search against all Arabidopsis proteins. Those candidate genes without a homolog in Arabidopsis (E-value of BLASTX hit >10E-6) were considered to be parts of TEs. This criterion bears the risk that genes which occur exclusively in grasses are missed. However, combined with the information that they are also not conserved in colinear positions, we considered this risk as small. One candidate gene from *Brachypodium* and six from rice were excluded as genes based on this criterion. Four of them were found in a region of *Osat_Lr34* with a high density of fragments of Mutator elements. It is likely that these genes are part of these elements.
- Third, TEs frequently "hijack" fragments of genes (Jiang *et al.*, 2004; Morgante *et al.*, 2005; Wicker *et al.*, 2003a). Typically, TEs carry only parts of genes (e.g. a few exons) and hardly ever complete CDS which can be used as an indication for a transposon-generated genic sequence. Indeed, three of the predicted *Brachypodium* candidate genes are only fragments which cover one, two and four exons, respectively. A

putative polygalacturonase gene (*PG-1*) from *Brachypodium* appears to be a processed pseudogene since it contains no introns in *Brachypodium* but has five in rice. Additionally, the gene is truncated at the 5' end and a poly-A stretch is found downstream of the CDS, suggesting an incomplete reverse transcription of an mRNA.

The Lr34 *loci from* Brachypodium *and rice have almost the same size but differ in gene density*

Excluding TE-related candidate genes, we estimate the number of "real" genes to be 66 in rice and 43 in *Brachypodium.* Of these, 61 and 38, respectively, encode proteins and five in each species are real or putative CNS. In total, the *Brachypodium* sequence contains 49,041 bp of annotated gene sequences (13.2% of the total sequence). This includes all CDS plus the five CNS regions. 30.5% of the sequence are annotated as repetitive DNA. The rice sequence contains more gene sequences (82,895 bp or 21.8%) and fewer repeats (9.9%). Consequently, *Osat_Lr34 has* a gene density of one gene every 5.8 kb whereas *Bsil_Lr34* has only one gene every 8.6 kb.

Of the 43 *Brachypodium* genes, 39 have orthologs in the colinear rice region. Among the conserved genes are the two tRNA genes *TRA* and *TRG* as well as the putative CNS regions *CR1, CR-2* and *CR-3*. Due to multiple gene duplications in rice, the 39 colinear *Brachypodium* genes correspond to 47 genes in rice (Figure 3). The two *ENO* genes in *Brachypodium* correspond to a cluster with seven members in rice and the single copy pectate lyase gene in *Brachypodium* (*BsPEC-1*) is present in three copies in rice. Interestingly, two of

the rice *PEC* genes are themselves part of an additional, smaller inversion (position 347'491-361'095) inside the large inversion.

Two additional duplications in rice resulted in the two partial genes *EP-2* and *LCM-2*. Curiously, *EP-1* and its 165 bp partially duplicated counterpart *EP-2* are separated by more than 40 kb and the segment in between them contains two non-colinear genes and several Mutator transposons (Figure 3, region α). The duplication *EP-1/EP-2* appears to be a result of the insertion of the large non-colinear DNA fragment and is reminiscent to the generation of target site duplications that are generated during the insertion of transposable elements into the genome. In contrast to rice, *Brachypodium* contains only one gene duplication, the partially duplicated *BsDRP-1* and *BsDRP-2* genes.

In *Brachypodium*, the four genes non-colinear with rice are interspersed with genes colinear to *Osat_Lr34*. In contrast, in the rice sequence most non-colinear genes are contained in four blocks of two to five genes (Figure 3, regions α,β,γ,δ) and only a few are interspersed, indicating the insertion or deletion of larger segments in one of the genomes. Interestingly, three non-colinear segments contain tandem arrays of related genes (regions α, δ and ε). Region ε contains a small cluster of three NBS-LRR resistance gene analogs (RGAs). The three genes are in the same transcriptional orientation and of the same evolutionary lineage (i.e. originated from duplications) as they are 60-70% identical at the DNA level. They are closely spaced within 20 kb. Interestingly, the *LCM-1* gene (which is conserved between rice and *Brachypodium*) is located between *RGA-1* and *RGA-2*, indicating that more than one rearrangement step (e.g. a single deletion) was necessary to reach the present situation. Regions α and δ contain two and five copies of F-box and transferase family genes, respectively.

In *Brachypodium*, the inverted segment is slightly larger (232 kb vs. 212 kb in rice) but contains fewer genes (21 vs. 36 in rice). The inverted region in *Brachypodium* is expanded by two large stretches that consist exclusively of repetitive elements (Figure 3).

No transposable elements are conserved in the two species which is in agreement with previous findings that most transposable elements in grasses inserted within the last few million years and that older elements are removed from the genome by various deletion mechanisms (SanMiguel *et al.*, 2002, Wicker *et al.*, 2003b).

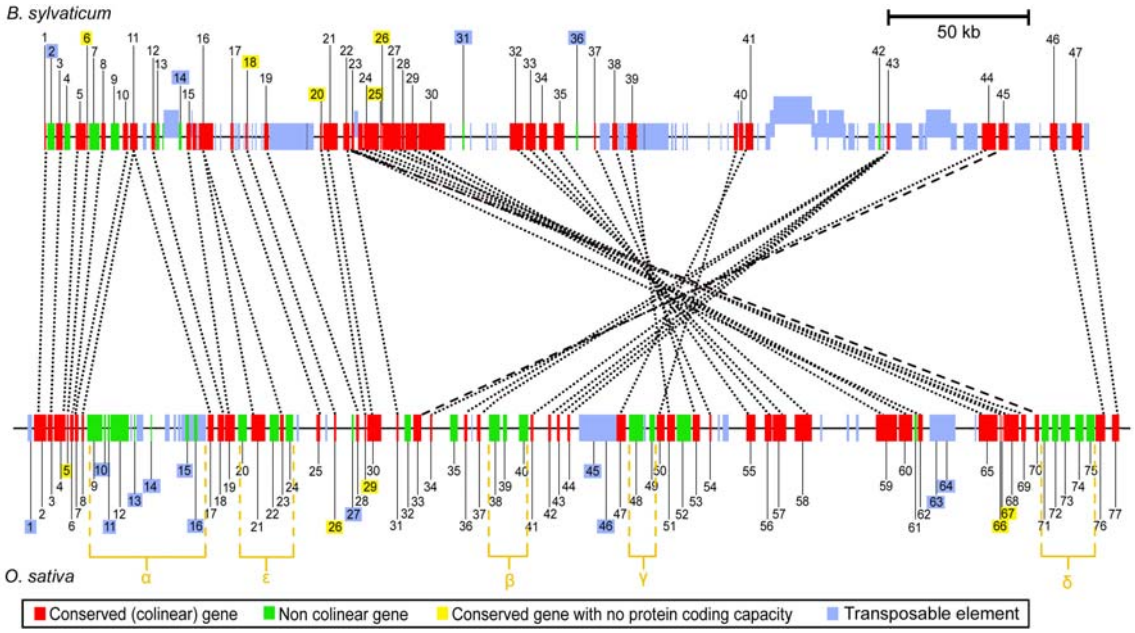

Figure 3. Comparison of the *Lr34* locus from *Brachypodium sylvaticum* (top) and rice (bottom). Regions specifically discussed in the text are indicated with Greek letters. Genes conserved between *Brachypodium* and rice are depicted as red boxes whereas non-conserved genes are depicted as green boxes. Orthologous genes are connected by black dotted lines. Transposable elements are indicated as blue boxes. Transposable elements that have inserted into others are raised to illustrate the nesting level. All candidate genes from both sequences are numbered. Numbers correspond to gene numbers in Supplementary tables 1 and 2. Genes with numbers highlighted in yellow are putative non-protein coding genes. Those highlighted in blue were classified as belonging to transposable elements.

Comparison of Brachypodium *and rice sequences can assist the annotation of both genomes*

Based on the observation that mainly CDS are conserved between *Brachypodium* and rice, we used DotPlot alignments to refine or modify the existing rice gene annotation. We compared all 47 protein-coding rice genes for which we had *Brachypodium* homologs (the 43 colinear and four non-colinear genes). In seven out of 47 cases, the predicted gene structure from rice had to be modified because either predicted CDS from rice did not correspond to conserved regions or splice donor and acceptor sites were not present at the expected positions in *Brachypodium* (Supplementary Table 2). For example, comparison of *ENO* genes showed that all members of the *ENO* family share a well-conserved CDS size and exon/intron structure, and allowed to deduce a more likely annotation for *OsENO-1* and *OsENO-2* (corresponding in TIGR annotation to *Os06g04930* and *Os06g04940*). The comparison also revealed that the divergent *OsENO-1* gene (*Os06g04930*) also belongs to the *ENO* cluster in rice.

For three of the seven re-annotated genes, Triticeae and rice ESTs were identified that helped to determine the precise position of exons and introns. The earlier predicted rice gene *Os06g04920* has 3 short exons and encodes a zinc finger motif protein. DotPlot analysis showed that *Os06g04920* is actually part of string of conserved motifs of more than 6 kb. We identified multiple Triticeae ESTs which cover parts of that region at ~80% sequence identity, indicating that *Os06g04920* is part of a larger gene with at least 5 exons (referred to as *ZNF-1*) which correspond to the conserved stretches in this region (Figure 4a). RiceGaas software predicted two genes in the flanking regions of *ZNF-1* where no sequence conservation between *Brachypodium* and rice was found but it failed to predict a gene in the conserved region (Figure 4a).

DotPlot also helped to modify the annotation of *Os06g05120* as it allowed the identification of two conserved regions in addition to the one that corresponds to the predicted gene *Os06g05120.* Interestingly, we found two ESTs from rice and one from wheat that indicate that *Os06g05120* can produce two alternative transcripts (Figure 4b). The second of two exons is the same in both transcripts whereas the first exon is different. The two alternative first exons are ~80% identical and apparently originated from an ancient duplication that predated the divergence of *Brachypodium* and rice (Figure 4b). Similarly, wheat ESTs helped develop two alternative models for a gene *Os06g05150*, one with 4 and one with 3 exons (Supplementary table 2).

In summary, 47 rice genes were analysed and the annotation of seven was modified based on comparison with their *Brachypodium* orthologs. In two cases, the location of the start codon was re-annotated. Three changes were made in the exon/intron structure and two involved the identification of alternative transcripts. If the region analysed is representative for the entire rice genome, one has to expect that about 15% of the rice gene annotations will be modified based on a completely sequenced *Brachypodium* genome.

After having included the modified annotations of the seven rice genes, we wanted to study the degree of conservation of the gene structure between *Brachypodium* and rice. In general, the exon/intron structure predicted for rice genes corresponded very well with the regions conserved between *Brachypodium* and rice. The 47 genes consist in total of 209 exons. Of the total 328 exon/intron boundaries, 318 were perfectly conserved in both species and all these 318 contain the consensus GT…AG motifs at the intron borders. Eight exon/intron boundaries were not found at expected positions and were predicted 3-21 bp away from it in *Brachypodium*. Interestingly, two boundaries are likely to have alternative splice donor or acceptor motifs: the third intron of the *Os06g05180* gene starts with a GG splice donor motif

which is conserved in *Brachypodium* and rice and for the ortholog of *Os06g4910* we found a wheat EST that suggests the 4th intron to have a GC splice acceptor motif in *Brachypodium*.

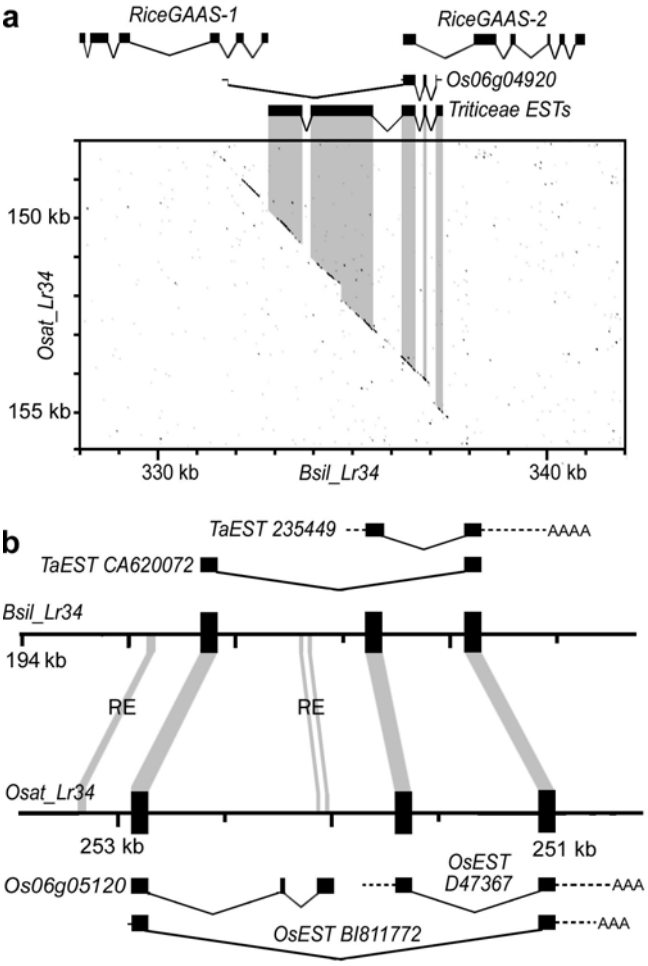

Figure 4. Identification of new putative genes through comparison of genomic sequences from *Brachypodium* and rice. **a.** Dotplot comparison of *Bsil_Lr34* (horizontal) and *Osat_Lr34* (vertical) suggests the presence of a gene with at least 6 exons. Exon/intron structure deduced by comparison with multiple Triticeae ESTs is indicated above the horizontal axis with exons depicted as solid bars and introns as bent lines. Grey areas illustrate how conserved regions correspond to putative exons. Positions of genes predicted by RiceGaas and the only gene annotated in this region by TIGR (*Os06g04920*) are indicated above the alignment. The prediced cDNA of *Os06g04920* contains a >5 kb intron but does not include most of the conserved motifs. **b.** Characterisation of the structure of rice gene *Os06g05120* and its ortholog from *Brachypodium*. Conserved sequences are indicated by grey areas connecting the maps of *Brachypodium* and rice. Regions conserved are exons (black boxes) as well as putative regulatory elements (RE). EST data from rice (prefix *OsEST*) and wheat (prefix *TaEST*) indicate that the gene encodes two alternative transcripts. Untranslated regions covered by ESTs are indicated by dotted lines and the position of poly-A tails is indicated by a stretch of A's. The predicted rice gene *Os06g05120* covers only one exon of the gene but has two predicted exons in regions that are not conserved between *Brachypodium* and rice.

In contrast to the highly conserved exon/intron structure, the 5' and 3' ends of the coding regions were more variable. In 8 genes, the position of the start codon deviated by 3-69 bp from the position expected from comparison with their predicted rice orthologs. Similarly, the stop codons of eight genes were 3-84 bp away from the positions expected from the comparison with their rice orthologs. This might either reflect problems with annotation or lower selection pressure on the N- and C-terminal regions of proteins.

The Lr34 *region in wheat is more colinear with* Brachypodium *than with rice but is greatly expanded and contains several non-colinear genes*

We also wanted to compare the characterised *Brachypodium* region with its orthologous locus in wheat. In a previous study, wheat ESTs that correspond to coding sequences of the rice *Lr34* orthologous region have been identified and used for genetic mapping of wheat chromosome 7D (Bossolini *et al.*, in press). Two of these markers (*BE493812* and *BF473324*) mapped on the orthologous region of the wheat chromosome 7A (Figure. 5). However, most of the wheat ESTs were monomorphic in the 'Arina' x 'Forno' mapping population (Paillard *et al.* 2003) for both the 7A and 7D chromosomes. Therefore, the ESTs were used to screen BAC libraries of hexaploid wheat (*T. aestivum*) and diploid *Ae. tauschii* to derive further markers. Two BACs from *Ae. tauschii* (HD007J19 and RI033N19) were identified using the EST probes *BJ220373* and *BE443044*, respectively and three BACs from *T. aestivum* (Ta940L4, Ta1466J2 and Ta470M18) where obtained by PCR screening with the previously published markers *GWM1220*, *STS77* and *GWM295*, respectively (Bossolini et al., in press).

To generate additional genetic markers and to investigate their gene content, the five BACs were low-pass shotgun sequenced to a 1 to 2-fold coverage. One shotgun clone (*SGC77*) from

BAC Ta1466J2 could be mapped as RFLP probe and SSR markers were derived from BACs RI033N19 (*SWM1*) and HD007J19 (*SWM5* and *SWM6*). *SWM1* and *SWM5* mapped on wheat chromosome 7D whereas *SWM6* and *SGC77* mapped on 7A (Figure. 5). Shotgun sequences were also used to anchor the five BACs to the *Brachypodium* and rice sequence (see below).

Integration of genetic mapping data from wheat with the physical maps of *Brachypodium* and rice revealed that *Brachypodium* and wheat share the same order of markers and that the inversion is only present in rice (Figure 5). The two *Ae. tauschii* BACs (HD007J19 and RI033N19) are part of two fingerprinted contigs with a cumulative size of almost 2 Mbp (Bossolini *et al.*, in press), indicating that the orthologous region in wheat is greatly expanded. Interestingly, the genetic distances in wheat are more than 10-fold larger than mapping distances in the integrated genetic map of rice (http://rgp.dna.affrc.go.jp/cgi-bin/giot/ine.pl, Figure 5).

Low-pass sequencing also showed that all five wheat BACs indeed contain genes that have homologs in the *Lr34* region in *Brachypodium* and rice (Figure 5). The five BACs contain a total of 19 candidate genes. Four of them, are of possible TE origin according to the criteria described above and were excluded from the gene content estimate. Two of these four are isolated exons of genes, indicating that they are part of a TE. For the two others, it is possible that they are simply not covered completely by the BACs as only one end of the gene is present in the low-pass BAC sequences. A total of 10 putative genes were identified that have orthologs in the *Brachypodium* sequence, 9 of which also have orthologs in rice. The remaining four gene candidates appear to be intact full-length genes based on comparison with Arabidopsis and rice genes. Thus, we estimated the five Triticeae BACs to contain 15 genes, 11 of which are conserved in *Brachypodium* and four are non-colinear genes (Figure 5, Supplementary Table 3). One gene from *Brachypodium* that was expected on BAC HD007J19

was not found, although the neighbouring up- and downstream genes in *Brachypodium* were present on HD007J19 (Figure 5).

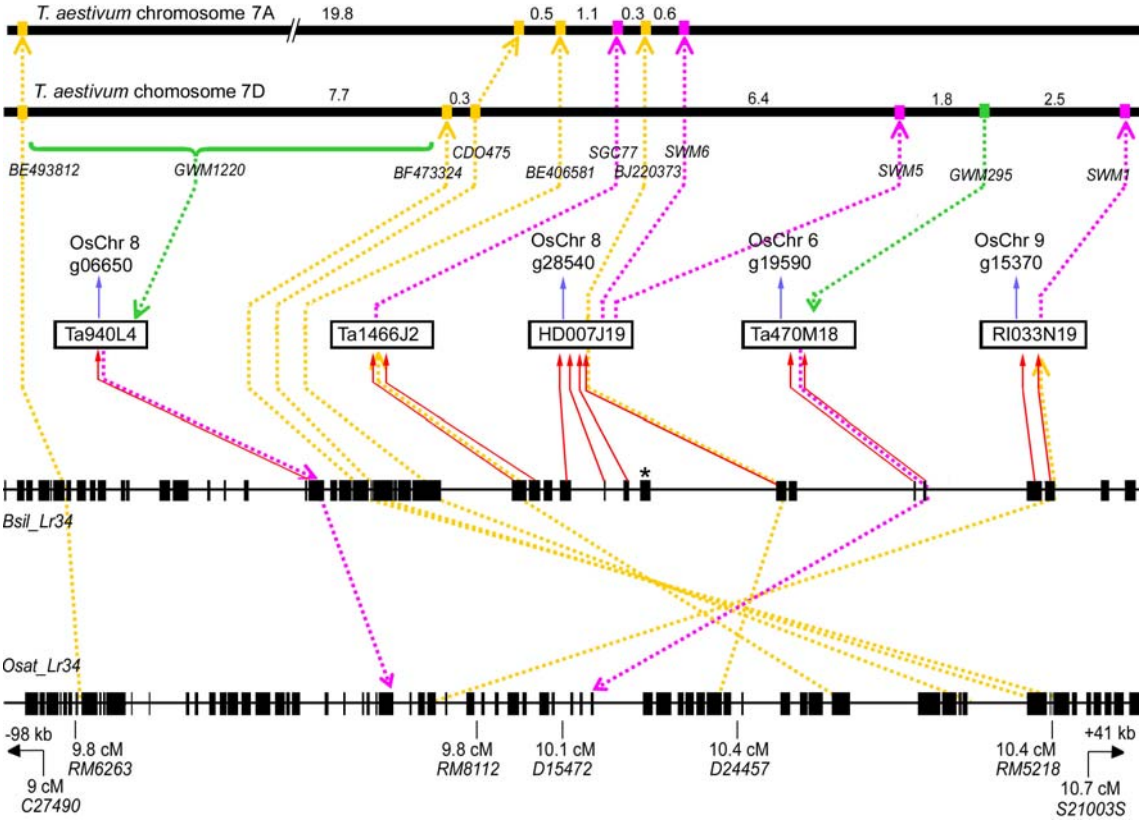

Figure 5. Comparison of the genetic map in wheat chromosomes 7D and 7A (top) with physical maps of the *Lr34* orthologous regions from *Brachypodium* (middle) and rice (bottom). Positions of genetic markers are indicated as coloured boxes on the wheat map and distances in cM are indicated between them. Marker positions on the integrated genetic map of rice are indicated underneath the rice map. Genes are indicated as black boxes in the physical maps. Wheat BACs anchored to the genetic and physical maps are indicated as boxes with BAC addresses. Dashed coloured lines connect corresponding locations of markers on the genetic and physical maps. Yellow dashed lines with arrows indicate positions of rice sequences that could be used to derive markers for genetic mapping in wheat and for the identification of wheat BACs. Markers derived from BACs that could later be anchored to the genetic or physical maps are indicated as purple dotted lines. Published markers from wheat that were used to identify wheat BACs are indicated in green dotted lines. Red solid arrows indicate homologs of *Brachypodium* genes found on wheat BACs. On four wheat BACs, genes were found whose putative rice orthologs (labelled with TIGR gene identifiers) map to different rice chromosomes (blue arrows). The asterisk indicates a *Brachypodium* gene whose ortholog is missing in the wheat BAC.

Gene duplications and deletions result in complex evolutionary relationships

Wheat BAC Ta470M18 contains two early nodulin (*ENO*) genes that are colinear with the two *ENO* genes from *Brachypodium* and seven from rice. Thus, we wanted to investigate whether different evolutionary lineages could be distinguished and whether certain lineages were deleted in some of the three species.

Hypothetical proteins were deduced for all 11 *ENO* genes from the three species and were used in a multiple sequence alignment using an Arabidopsis *ENO* homolog (*At5g25940*) as an outgroup. Phylogenetic analysis revealed complex relationships between orthologs and paralogs and allowed to classify the *ENO* genes into three main evolutionary lineages (I, II, and III, Figure 6). Lineage I contains six of the rice genes (*OsENO-2* through *OsENO-7*) as well as the *Brachypodium BsENO-2* gene. Thus, the six rice genes are all orthologous to *BsENO-2* (Figure 6). Lineage II contains the two *ENO* genes from wheat and *BsENO-1* whereas lineage III is exclusive to rice and contains only the most divergent gene *OsENO-1* (Figure 6). Based on this analysis, we propose a model in which the common ancestor of the three species contained members of all three lineages (Figure 6c). after the separation of rice and *Brachypodium*/wheat, lineage II was deleted in rice whereas lineage III was deleted in the ancestor of *Brachypodium* and wheat. After, the evolutionary separation of *Brachypodium* and wheat, lineage I was deleted in wheat.

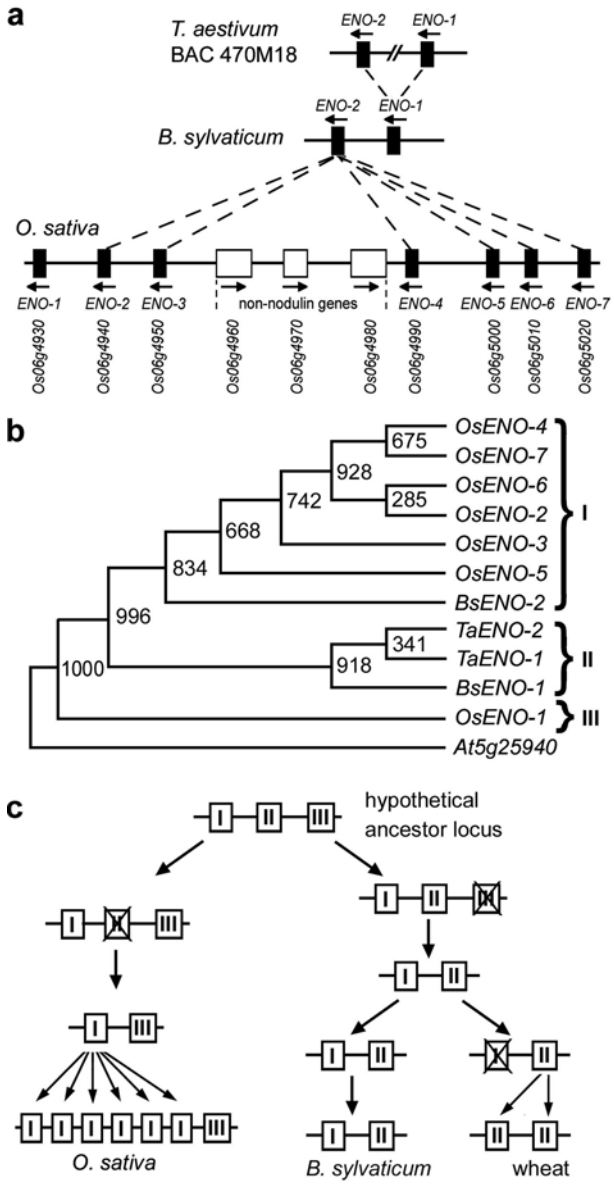

Figure 6. Analysis of early nodulin (*ENO*) genes in *Brachypodium*, rice and wheat (*T. aestivum*). **a.** Sequence organisation of the *ENO* loci in the three species. *ENO* genes are indicated as black boxes. The cluster in rice is interrupted by a segment containing three genes of different families (white boxes). Transcriptional orientations are indicated by horizontal arrows. **b.** Phylogenetic analysis of predicted protein sequences of *ENO* genes shows three major evolutionary lineages (I, II and III). Species are indicated with the prefixes *Bs* (*B. sylvaticum*), *Os* (*O. sativa*) and *Ta* (*T. aestivum*). The Arabidopsis protein which gave the best BLASTP hit (*At5g25940*) was used as outgroup to root the tree. **c.** Model of the evolution of *ENO* genes. The hypothetical ancestor locus contains three genes which correspond to the three major evolutionary lineages. Deletion events are indicated by crossed out genes.

Brachypodium *and Triticeae diverged approximately 35 million years ago*

To estimate the divergence time of *Brachypodium* and Triticeae, CDS from the nine genes which are conserved in *Bsil_Lr34* and *Osat_Lr34* and were also found in the low-pass sequenced BAC clones from wheat were compared by multiple alignments. Divergence time can only be estimated under the assumption that the compared genes are evolving at a constant rate in all three species (i.e. have a molecular clock). Therefore, only those sequences were used where *Brachypodium* and wheat were more closely related to each other than to rice and both showed approximately the same level of DNA sequence divergence form rice (i.e. divergence of *Brachypodium* and Triticeae sequences from rice was not significantly different at the $p<0.05$ level). This eliminates the possibility that a gene that evolves at a different rate (or degenerated pseudogene) is used which would distort evolutionary distances.

Five of the nine genes tested met these criteria and a rough estimate of divergence time was obtained by simply comparing the ratio of sequence divergence between the three species for the five genes under the assumption that the *Brachypodium*/Triticeae lineage diverged from rice approximately 50 MYA (Paterson *et al.*, 2004). The resulting estimates for the divergence time of *Brachypodium* and Triticeae lie between 29.9 and 39.7 million years (Table 1).

To broaden the sample, all CDS of genes conserved between *Brachypodium* and rice were used as queries in BLASTN searches against all Triticeae ESTs. If the entire CDS was covered by multiple Triticeae ESTs, a consensus sequence was produced. In cases where ESTs covered only parts of the CDS, as many as possible overlapping ESTs were identified and connected manually in order to cover the largest possible portion of the respective gene. We did not differentiate between individual Triticeae species, assuming that they are all equally distantly related to *Brachypodium*. For 32 genes, sufficient EST sequences were

found to allow a multiple alignment of *Brachypodium*, rice and Triticeae sequences and 18 met the criteria that *Brachypodium* and wheat sequences are equally divergent from rice (as described above).

Combining the data from genomic and EST sequences, 23 divergence time estimates were made (5 from genomic/BAC DNA and 18 from ESTs). For two genes (*ST-1* and *HP-4*), data were available from both BAC DNA and ESTs. *Brachypodium*/Triticeae sequence pairs were 75.6% - 91.7% identical at the DNA level, whereas the rice/*Brachypodium* and rice/Triticeae pairs ranged from 67.3% to 88.3% sequence identity. The estimated divergence times from genomic CDS and ESTs ranged from 20.2 to 39.8 MYA, with an average of 34.2 MYA (Table 1). However, 16 of the 23 values are in the relatively narrow range of 34.7 – 39.9 MYA. For *ST-1* and *HP-4*, 100% wheat EST matches were found and divergence time estimated from genomic (BAC) DNA and from ESTs are very similar.

For three genes, the estimates suggested much more recent divergence times of 20.2 to 27.3 million years, respectively (Table 1). In the case of the *ENO-1* gene, we know that paralogs were compared (see above) which can explain the strong discrepancy. The same may be true for the other genes whose estimated divergence times differ much from the bulk. Indeed, for the *KIN* gene, the estimated divergence time based on genomic DNA from the BAC clone Ta940L4, is 38.4 MYA whereas the one based on Triticeae ESTs is 27.3 MYA. The *KIN* gene found on the wheat BAC has no 100% match in wheat EST databases, indicating that the identified wheat ESTs originate from a paralogous gene elsewhere in the genome. The same was found for the *PKD* and *EP-4* genes. Thus, we are inclined to ignore the estimates that are spread between 20 and 34 MYA and consider those between 34.7 and 39.9 MYA as being close to the actual value.

Table 1. Estimates of divergence time of *Brachypodium* and wheat based on comparison of 23 genes. Note that for two Triticeae genes, sequence data from both genomic DNA and ESTs was available.

Gene	divergence MYA
PKD-1[a]	20.2
ZP-1[a]	23.8
ENO-1[b]	27.3
KIN-1[a]	29.6
PKD-1[b]	29.7
GBP-1[a]	31.8
PAB-1[a]	32.7
MSR-1[a]	34.7
LCM-1[a]	35.0
EP-1[a]	35.1
EP-4[b]	35.2
ZCP[a]	35.5
ST-1[b]	35.5
PGI-1[a]	35.9
DRP-1[a]	35.9
HP-4[a]	36.6
ST-1[a]	37.6
KIN[b]	38.4
EP-4[a]	38.8
Hox-1[a]	38.9
OxR[a]	39.8
HP-4[b]	39.8
AAT[a]	39.9

[a]Estimates based on Triticeae ESTs with from unknown genomic loci.

[b]Estimates based on genomic CDS from wheat BAC clones from *Lr34* orthologous genomic loci.

6.5 Discussion

The presented study gives an insight into a 371 kb segment of the *Brachypodium* genome and its evolutionary relationship to orthologous regions from the two crops species rice and wheat. The three species show good macro-colinearity but distinct genomic rearrangements were

found at the micro-colinearity level. We showed that comparative annotation of genomic sequences from *Brachypodium* and rice can improve annotation and helps identify novel, potentially functional, DNA motifs. We estimated *Brachypodium* to have diverged from Triticeae about 34-40 MYA and showed it to be clearly more closely related to the Triticeae than rice in many aspects. However, our data also show the *Brachypodium* genome still differs considerably from Triticeae genomes in gene order and gene content.

Comparison of two small grass genomes provides quantitative data on micro-colinearity and sequence conservation

The rice genome is the only complete grass genome sequence available to date and it is also the smallest grass genome that is well studied. Other crop genomes for which sequence data are available are all considerably larger (e.g. wheat, maize, *Sorghum*) and comparative genomics in grasses is currently restricted by two factors: First, comparative analysis is only possible between rice and species with larger genomes and, second, the compared chromosomal regions were very limited in size because sequencing BAC clones from large and repetitive genomes is very time-consuming and yields only a few gene sequences. Thus, most previous comparisons were restricted to regions that contain only a few (Chantret *et al.*, 2004; Bennetzen and Ma, 2003; Dubcosvky *et al.*, 2001; Guyot *et al.*, 2004; Song *et al.*, 2002; Wicker *et al.*, 2003b) and hardly ever more than ten genes that are conserved between species (Ma *et al.* 2005; Swiganova *et al.* 2005). In contrast, our study allowed comparison of a large genomic region between two small grass genomes containing dozens of conserved genes. Therefore, the compared sequences also allow some quantitative analysis.

Only few previous studies provided data on conservation of overall gene structure such as exon/intron boundaries. Analysis of genes from the *waxy* and *vrn-1* loci showed a that exon/intron structure is generally conserved between species but several exceptions were found and no prediction as to the extent of conservation could be made (Bennetzen and Ma, 2003; Dubcovsky *et al.*, 2001). We analysed in detail 47 rice genes and their homologs from *Brachypodium* and found that most genes have a highly conserved structure. Therefore, it can be expected that at least 95% of exon/intron boundaries will be found conserved between grass species.

Interestingly, the rice *Lr34* locus has considerably higher gene content than its counterpart from *Brachypodium*, although both regions are very similar in size. In contrast, only four genes present in the *Brachypodium* sequence are not conserved in rice. Thus, almost all *Brachypodium* genes are perfectly colinear in rice. We also observed less of a "mosaic" conservation of genes between the two small genomes than previous studies found comparing rice with larger grass genomes (Song *et al.*, 2002;Swiganova *et al.*, 2005), since only few single non-colinear genes were found interspersed with colinear genes. Curiously, it appears that in rice, multiple large segments containing several genes have been inserted (or removed from the corresponding regions in *Brachypodium*), suggesting frequent movement of large segments.

Gene density in the rice sequence is more or less even across the whole region analysed, whereas in *Brachypodium*, the first 210 kb of the contig contain most genes and gene density is almost identical to that in rice (one gene every 6 kb). The rest of the sequence is mainly comprised of repetitive DNA and has a gene density of one gene every 20 kb which is reminiscent of the situation found in some regions of the Triticeae genomes (SanMiguel *et al.*, 2002; Wicker *et al.*, 2001). It can therefore be expected that the *Brachypodium* genome

contains extensive regions in the range of hundreds of kb that are comprised almost exclusively of genic sequences but also may contain highly repetitive regions where sequencing and analysis will be almost equally difficult as in large grass genomes.

One of the most striking findings of this study are several conserved putative genes that apparently do not have any protein coding capacity. Only the two genes for tRNA genes could be assigned a function whereas the function of the others is unknown. The presence of such conserved noncoding sequences in plants was reported previously (Kaplinsky *et al.*, 2002) but only a few isolated genes have been studied. Surprisingly, five of the 39 conserved genic sequences in *Brachypodium* are CNS, indicating that a considerable fraction of functional sequences in the plant genome may not encode any proteins. Especially for the conserved region *CR-1* it is not clear if it represents an unknown gene with several exons or a series of independent motifs that are conserved in colinear positions. If one assumes CR-1 to be just one gene, extrapolation of the numbers suggest that 10-15% of the functional sequences in the rice or *Brachypodium* genomes might be CNS.

Kaplinsky *et al.* (2002) also described several short CNS that are in the size range of only a few dozen bp. We identified several such short sequence motifs associated with most genes, suggesting them to be conserved regulatory elements. These findings reveal the potential of comparative genomics in two small grass genomes to identify completely new structures with potentially important functions. Future larger-scale comparative analysis will potentially reveal a large number of CNS for subsequent functional analysis.

Comparative annotation identified new genes and improves genome annotation

For our analysis, we used very stringent parameters for gene annotation which were in part based on suggestions by Bennetzen *et al.* (2004) in order to minimise the number of TE sequences that are annotated as genes. Previous studies showed that gene content is often over-estimated due to the TE sequences being annotated as genes (Bennetzen *et al.*, 2004; Morgante *et al.*, 2005; Wicker *et al.*, 2003a). Our initial annotation that included all candidate genes predicted 77 genes for the rice and 47 gene for the *Brachypodium* sequence. Removal of all genes that are of possible TE origin dropped the numbers to 66 and 43, respectively, increasing the percentage of genes conserved between the two species accordingly. Therefore, the apparent degree of conservation or colinearity of two compared genomes depends, in part, on the annotation of the compared sequences.

Even the presence of ESTs does not prove gene functionality because some gene fragments found in Helitrons are also expressed (Bennetzen *et al.*, 2004, Morgante *et al.*, 2005). Our data suggest that the best indication that a sequence is indeed a real gene is its conservation in colinear positions in different species and every candidate gene that does a not have a homolog in a colinear position should be considered to be of possible TE origin. Occasionally, conserved motifs may appear within repetitive DNA (e.g. the reverse transcriptase domain of retrotransposons) but only very rarely they will also be found in colinear positions so that the could be mistaken for genic sequences.

Comparison of *Brachypodium* and rice sequences also identified several instances where the existing rice annotation could be modified to match regions conserved between the two species. Although the vast majority of rice genes seem to be predicted very reliably, we were able to unravel some important differences such as the identification of alternative transcripts.

Our data indicate that up to 15% of the predicted rice genes may require some sort of modification. In most cases, however, we expect these to be very minor changes such as slight shifts in positions of start or stop condons or splice sites. A recent study showed that ~21% of the genes annotated in Arabidopsis were modified using random shotgun clones from *Brassica* (Katari *et al.*, 2005). Considering that the Arabidopsis genome is likely the best-annotated plant genome, these numbers emphasize the importance of such comparative annotation.

Brachypodium *as bridge genome between rice and Triticeae?*

The term "bridge genome" reflects the assumption that *Brachypodium* is a species with the gene order of Triticeae species but the genome size of rice. In the ideal case, information on its complete genomic sequence would make sequencing of the more than ten times larger Triticeae genomes unnecessary. Our data show a perfect colinearity of genetic markers between *Brachypodium* and wheat. The large inversion that is found in rice but absent in *Brachypodium* and wheat places *Brachypodium* clearly closer to wheat and indicates that *Brachypodium* may be a better anchor genome for map-based cloning of wheat genes. These findings are in agreement with previous studies which described a high degree of conservation of genetic markers between *Brachypodium* and Triticeae (Griffiths *et al.*, 2006; Hasterok *et al.*, 2006).

The *Brachypodium* genome sequence could play a significant role in the study of evolutionary relationships between grasses. The example of the *ENO* gene family shows that gene loss of different paralogs in rice and Triticeae can obscure the true evolutionary relationships within gene families. Previous studies showed that considerable portions of grass genomes may

contain amplified genes (Song *et al.*, 2002) and that true evolutionary relationships can only be revealed if more than two species are compared (Feuillet *et al.*, 2001). Thus, the availability of a second small genome will be a very valuable resource to investigate orthologous and/or paralogous relationships of genes from Triticeae, maize, *Brachypodium* and rice.

Our data also show that considerable differences are found between *Brachypodium* and wheat at the micro-colinearity level. Data from wheat BAC clones show that the *Lr34* region in wheat is greatly expanded and probably has a size of several Mbp. Previous studies showed that the difference in size is usually caused by large amounts of intergenic transposable element rather than a drastic difference in gene content (Bennetzen and Ma, 2003; Swigonova *et al.*, 2005). However, it is striking that the genetic distance covered by the described markers in wheat is more than 10-fold larger than in rice although the total length of the integrated genetic map of rice (http://rgp.dna.affrc.go.jp/cgi-bin/giot/ine.pl) is with 1500 cM larger than the genetic map of the wheat A and D genomes (1131 and 1036 cM, respectively, Paillard *et al.*, 2003). This could, on one hand simply be a statistical fluctuation as recombination rates tend to change considerably along chromosomes. On the other hand, it could indicate that wheat may contain introgressions of large chromosomal segments in this region since high recombination rates are often correlated with high gene content. However, we would have to postulate multiple large insertions in the wheat genome since genetic distance between most wheat markers are expanded compared to rice. Such a scenario is not excluded especially considering our finding of multiple non-colinear segments in rice compared to *Brachypodium*.

Only about two-thirds of the genes from five wheat BACs from the *Lr34* region were found to be colinear with *Brachypodium* and one *Brachypodium* gene had no ortholog in the wheat

BAC where it was expected. These data indicate that great caution must be used if *Brachypodium* should serve as an anchor genome for map-based cloning in Triticeae.

Finally, one should consider that *Brachypodium* has diverged prior to the radiation of the Pooideae which include many forage grasses and temperate cereals and are characterised by large genomes (Catalan and Olmstead, 2000; Draper *et al.*, 2001). According to our analysis, this divergence took place approximately 35-40 MYA. Thus, *Brachypodium* as a model is about 10-15 million years "closer" to Triticeae than rice. This is still a considerable phylogenetic distance to the Triticeae which diverged only about 10-14 MYA (Wolfe *et al.*, 1989). Thus, a wide range of genomic changes must be expected between *Brachypodium* and Triticeae. For example, the enormous difference in genome size between *Brachypodium* and Triticeae is an obvious manifestation of the great changes the genomes of the two evolutionary lineages have undergone.

We conclude that the complete genome sequence of *Brachypodium* will be of enormous relevance for comparative genomics and gene annotation but it will possibly not be able to substitute for the large-scale analysis of Triticeae genomes.

Acknowledgments:

This work was supported by the Swiss National Science Foundation (SNF) grant 3100AV-105620. We would like to thank Dr. Edith Schlagenhauf for her support in bioinformatics and Dr. Olin Anderson and Dr. Curt Crossman (USDA-ARS, Albany, CA, USA) for providing wheat ESTs. We acknowledge contribution of the Biotechnology and Biological Sciences Research Council as well as the John Innes Center (UK) for providing the BAC library and BAC filters of *Brachypodium sylvaticum*.

Chapter 7

GENERAL DISCUSSION

7.1 A diagnostic marker to integrate the *Lr34* resistance in wheat breeding lines

7.2 Comparative genomics to characterize small regions in wheat

7.3 Outlook

7. GENERAL DISCUSSION

7.1 A diagnostic marker to integrate the *Lr34* resistance in wheat breeding lines

The main reason for crop disease is related to the very purpose of agriculture: the high density cultivation of homogeneous genotypes. This aspect is so widespread to all cropping systems that it is commonly referred to as 'the original sin of agriculture'. This unnatural situation exerts a dramatic evolution pressure on pathogen populations, leading to rapid fluctuations in the population of pathogen races and rapid adaptation to the pool of resistance genes present in the crop. This has been commonly observed in breeding programs for leaf and stripe rust resistance. Wheat cultivars having one or few vertical resistance genes became typically susceptible after a few years of extensive cultivation (Singh et al. 2005). Breeding for quantitative, horizontal resistance is a promising approach to achieve durable resistance in commercial germplasm. Vertical resistance genes in wheat have been widely deployed to limit pathogen epidemics in the field. They mostly interact with specific pathogen races according to the gene-for-gene model and unfortunately they are not durable. Horizontal defence genes like *Lr34/Yr18* and *Lr46/Yr29* do not confer a high degree of resistance to rusts, but they have been proven to slow down significantly the disease development (Niks and Rubiales 2002). A breeding program directed at pyramiding horizontal and vertical resistances can achieve a level of resistance close to immunity, with the advantage that they are not quickly overcome by some pathogen races (Singh et al. 2005). Until now, pyramiding of resistances was achieved with traditional breeding. Unfortunately, for some important adult plant resistance genes like *Lr34,* detection of the resistant phenotype is only possible after

flowering. This delays effective crossing of one generation, and doubles the efforts for breeding. Our new diagnostic marker SWM10 predicts the presence of *Lr34/Yr18* independently of the plant growth stage or environmental conditions and is extremely valuable to introgress resistance into susceptible germplasm.

7.2 Comparative genomics to characterize small regions in wheat

Since several years the rice genome sequence has been used to improve our understanding of the largely unexplored genomes of cereals like wheat, barley, maize and sorghum. While a genome sequencing initiative has been started for maize and sorghum, for none of the Triticeae species this is planned in the near future, except for the 3B chromosome of hexaploid wheat. At the moment, rice provides the only model genome for the Triticeae. However, the nature of mosaic conservation observed in many regions between wheat and rice limits the use of the rice genome sequence for positional cloning in wheat (Guyot et al. 2004). This is especially true for resistance genes that normally do not conserve synteny (Leister et al. 1998). Recently, there is an increased interest in the small grass genome *Brachypodium*. This temperate grass has a short life cycle, has a small genome (similar size as the rice genome), is easy to transform and represents a good model for functional genomic studies (Hasterock et al. 2001). Its genome sequence will soon be available (www.brachypodium.org). However, this thesis reveals that this model genome also has its limits when used to characterize orthologous regions in wheat. We have found that gene order in the studied region is more similar to wheat than to rice. Despite this, none of the resistance gene analogues observed in wheat did maintain synteny with *Brachypodium*. This was observed also for clusters of similar genes or gene families. If they originate from duplications predating the separation of the two species, they follow divergent evolutionary pathways, they

possibly evolve distinct functions and often they are reshuffled in the genome. More and more genomic sequence is becoming available from grasses with different time frames of independent evolution, giving an insight in the mechanisms resulting in genomic changes. The monocots (e.g. rice) and the dicots (e.g. *Arabidopsis*) diverged 140-200 MYA (Million Years Ago) (Wolfe et al. 1989): between these two plant species the majoritity of genes are conserved only at the protein level and they do not maintain syntenic position. Rice has similar divergence time with the Panicoideae (e.g. maize or sorghum, 50MYA, White and Doebley 1999) and the Triticeae (e.g. wheat or barley, 46 MYA, Wolfe et al. 1989). A collinear position between rice and Triticeae is maintained for half of the genes and there is similarity also at the nucleotide level; intergenic DNA is not conserved. Comparisons of maize and sorghum (which diverged 16.5-9 MYA, Gaut and Doebley 1997) revealed a higher degree of gene sequence conservation and synteny, but only after having taken into account that most of maize genes are pseudogenes shuffled by helitrons (Morgante et al. 2005), whereas intergenic DNA is not conserved. Similarly, within the Triticeae species (divergence occured 13 MYA) good synteny is observed for genes, whereas intergenic DNA differs. Some degree of similarity also in the intergenic DNA was described between the two rice subspecies *japonica* and *indica,* which diverged only 0.44 MYA (Ma and Bennetzen 2005). This highlights the rapid evolution rate of intergenic DNA. Concerning wheat, its three genomes share a degree of similarity comparable with other Triticeae species, whereas each subgenome is clearly closely related to its wild ancestor genome, including the intergenic DNA (Chantret et al. 2005). Since the introgression of the D-genome in wheat occurred 8,000 years ago, a degree of conservation close to identity was found between the wheat D-genome and the *Ae. tauschii* genome. Perfect sequence conservation is extended to intergenic regions as demonstrated by high transferability of genomic SSRs between *Ae. tauschii* and wheat (Guyomarc'h et al. 2002). Consequently, the partially fingerprinted BAC library of *Ae. tauschii* was extremely useful to saturate with more markers an important region of the wheat

genome. A more complete physical map of wheat or one of its progenitors would provide an excellent tool for wheat genomics and make the use of model genomes obsolete.

7.3 Outlook

Microsatellite markers specifically tagging a small region of the large wheat genome were developed using rice as a model genome and a fingerprinted BAC library of *Ae tauschii* to bridge the phylogenetic distance between rice and wheat. The information provided by the orthologous sequence of the *Brachypodium* genome has defined with more details the borders of the inversion detected between rice and wheat and allowed to focus marker development on a better defined region of ~15 kb in *Brachypodium* and rice. Besides the described inversion, gene content in this region is conserved between these two species. The same region in wheat is still unexplored. Information about size and gene content are to be gained. Synteny information indicate that a BAC contig of ~800 kb from *Ae. tauschii* is located within the *Lr34* genetic interval defined by the markers *Xgwm1220* and *Xsfr.BF473324*. However, this contigous DNA fragment does not contain the two *Lr34* flanking markers. Anchoring the BAC contig to the genetic map can be achieved by chromosome walking at both sides of the BAC contig. Unfortunately, the high amount of repetitive DNA and polyploidy often impede chromosome walking in wheat. In the future, more molecular resources will become available to characterize this region. More wheat chromosome and subchromosome BAC libraries will be produced by Dr. Jaroslav Dolezel (Laboratory of Molecular Cytogenetics and Cytometry, Czech Republic) within the frame of the international wheat genome sequencing consortium (IWGSC). Specifically, recently CSIRO and ASPFG (Australia) demonstrated interest in setting up a large project for physically mapping all group 7 chromosomes (7A, 7B, 7D). At present, the lack of a contiguous DNA sequence containing

Lr34 and anchored to the genetic map leverages the need to develop more molecular markers for this region. In the long term, this will enable map-based isolation of this agronomically important resistance locus.

8. REFERENCES

Abajian C (2003) Sputnik: C program to search DNA sequence for microsatellite repeats. http://espressosoftware.com/pages/sputnik.jsp

Akhunov ED, Goodyear AW, Geng S, Qi LL, Echalier B, Gill BS, Miftahudin, Gustafson JP, Lazo G, Chao SM, Anderson OD, Linkiewicz AM, Dubcovsky J, La Rota M, Sorrells ME, Zhang DS, Nguyen HT, Kalavacharla V, Hossain K, Kianian SF, Peng JH, Lapitan NLV, Gonzalez-Hernandeiz JL, Anderson JA, Choi DW, Close TJ, Dilbirligi M, Gill KS, Walker-Simmons MK, Steber C, McGuire PE, Qualset CO, Dvorak J (2003) The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Res 13:753-763

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-3402

Ayala L, Henry M, van Ginkel M, Singh R, Keller B, Khairallah M (2002) Identification of QTLs for BYDV tolerance in bread wheat. Euphytica 128:249-259

Bennett MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. Phil Trans R Soc Lond B, 274:227-274

Bennetzen JL, Ma J (2003) The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis. Current Opin Plant Biol 6:128-133

Bennetzen JL, Ramakrishna W (2002) Numerous small rearrangements of gene content, order and orientation differentiate grass genomes. Plant Mol Biol 48:821-827

Bennetzen, JL (2000) Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. Plant Cell, 12:1021-1030

Bennetzen JL, Coleman C, Liu R, Ma J, Ramakrishna W (2004) Consistent over-estimation of gene number in complex plant genomes. Curr Opin Plant Biol 7:732-736

Bossolini E, Krattinger SG, Keller B (2006) Development of SSR markers specific for the Lr34 resistance region of wheat using sequence information from rice and Aegilops tauschii. Theor Appl Genet, in press.

Brueggeman R, Rostoks N, Kudrna D, Kilian A, Han F, Chen J, Druka A, Steffenson B, Kleinhofs A (2002) The barley stem rust resistance gene Rpg1 is a novel disease resistance gene with homology to receptor kinases. Proc Natl Acad Sci USA 99:9328-9333

Brunner S, Keller B, Feuillet C (2003) A large rearrangement involving genes and low-copy DNA interrupts the microcollinearity between rice and barley at the *Rph7* locus. Genetics 164:673-683

Cadalen T, Boeuf C, Bernard S and Bernard M (1997) An intervarietal molecular marker map in *Triticum aestivum* L. em Thell and comparison with a map from a wide cross. Theor Appl Genet 94:367-377

Catalan P, Olmstead RG (2000) Phylogenetic reconstruction of the genus *Brachypodium* P. Beauv. (Poaceae) from combined sequences of chloroplast *ndhF* gene and nuclear ITS. Plant Syst Evol, 220:1-19

Chantret N, Salse J, Sabot F, Rahman S, Bellec A, Laubin B, Dubois I, Dossat C, Sourdille P, Joudrier P, Gautier MF, Cattolico L, Beckert M, Aubourg S, Weissenbach J, Caboche M, Bernard M, Leroy P, Chalhoub B (2005) Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum and aegilops*). Plant Cell 17:1033-1045

Chantret N, Pavoine MT and Doussinault G (1999) The race-specific resistance gene to powdery mildew, *MlRE*, has a residual effect on adult plant resistance of winter wheat line RE714. Phytopathology 89:533-539

Cregan PB, Mudge J, Fickus EW, Marek LF, Danesh D, Denny R, Shoemaker RC, Matthews BF, Jarvik T, Young ND (1999) Targeted isolation of simple sequence repeat markers through the use of bacterial artificial chromosomes. Theor Appl Genet 98:919-928

Draper J, Mur LAJ, Jenkins G, Ghosh-Biswas GC, Bablak, P, Hasterok R, Routledge APM (2001) *Brachypodium distachyon*. A new model system for functional genomics in grasses. Plant Physiol, 127, 1539-1555.

Dubcovsky J (2004) Marker-assisted selection in public breeding programs: The wheat experience. Crop Sci 44:1895-1898

Dubcovsky J, Ramakrishna W, SanMiguel PJ, Busso CS, Yan L, Shiloff BA, Bennetzen JL (2001) Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. Plant Physiol 125:1342-1353

Dyck PL (1991) Genetics of adult-plant leaf rust resistance in Chinese Spring and sturdy wheats. Crop Sci 31:309-311

Dyck PL, Samborski DJ, Martens JW (1985) Inheritance of resistance to leaf rust and stem rust in the wheat cultivar Glenlea. Can J Plant Path 7:351-354

Endo TR, Gill BS (1996) The deletion stocks of common wheat. J Hered 87:295-307

Ewing B, Hillier L, Wendl MC and Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res, 8,175-185.

Faris J D, Fellers JP, Brooks SS, Gill BS (2003) A Bacterial Artificial Chromosome contig spanning the major domestication locus *Q* in wheat and identification of a candidate gene. Genetics 164:311-321

Feng Q, Zhang YJ, Hao P, Wang SY, Fu G, Huang YC, Li Y, Zhu JJ, Liu YL, Hu X, Jia PX, Zhang Y, Zhao Q, Ying K, Yu SL, Tang, YS, Weng, QJ, Zhang L, Lu Y, Mu J, Lu YQ, Zhang LS, Yu Z, Fan DL, Liu XH, Lu TT, Li C, Wu YR, Sun TG, Lei HY, Li T, Hu H, Guan JP, Wu M, Zhang RQ, Zhou B, Chen ZH, Chen L, Jin ZQ, Wang R, Yin HF, Cai Z, Ren SX, Lv G, Gu WY, Zhu GF, Tu YF, Jia J, Chen J, Kang H, Chen XY, Shao CY, Sun Y, Hu QP, Zhang XL, Zhang W, Wang LJ, Ding CW, Sheng HH, Gu, JL, Chen ST, Ni L, Zhu FH, Chen W, Lan LF, Lai Y, Cheng ZK, Gu MH, Jiang JM, Li JY, Hong GF, Xue YB, Han B (2002). Sequence and analysis of rice chromosome 4. Nature 420:316-320

Feuillet C, Keller B (2002) Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution. Ann Bot 89:3-10

Feuillet C, Penger A, Gellner K, Mast A, Keller B (2001) Molecular evolution of receptor-like kinase genes in hexaploid wheat. Independent evolution of orthologs after polyploidization and mechanisms of local rearrangements at paralogous loci. Plant Physiol 125:1304-1313

Feuillet C, Travella S, Stein N, Albar L, Nublat A, Keller B (2003). Map-based isolation of the leaf rust disease resistance gene Lr10 from the hexaploid wheat (*Triticum aestivum* L.) genome. Proc Natl Acad Sci USA 100:15253-15258

Foote T, Griffiths S, Allouis S and Moore G (2004) Construction and analysis of a BAC library in the grass *Brachypodium sylvaticum* and its use as a tool to bridge the gap between rice and wheat in elucidating gene content. Funct Integr Genomics 4:26-33

Foote T, Roberts M, Kurata N, Sasaki T, Moore G (1997) Detailed comparative mapping of cereal chromosome regions corresponding to the *Ph1* locus in wheat. Genetics 147:801-807

Gale MD, Atkinson MD, Chinoy CN, Harcourt RL, Liu J, Li QY, DK M (1995) Genetic maps of hexaploid wheat. In: Li ZS, Xin ZY (eds) Proc 8th Int Wheat Genet Symp, Agric Scientech Press, Beijing, pp 29–40

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci USA 95:1971-1974

Gallego F, Feuillet C, Messmer M, Penger A, Graner A, Yano M, Sasaki T, Keller B (1998) Comparative mapping of the two wheat leaf rust resistance loci *Lr1* and *Lr10* in rice and barley. Genome 41:328-336

Gaut BS and Doebley JF (1997) DNA sequence evidence for the segmenal allotetraploid origin of maize. Proc Natl Acad Sci USA 94:6809-6814

German, S. E. and Kolmer, J. A. 1992. Effect of gene Lr34 in the enhancement of resistance to leaf rust of wheat. Theor. Appl. Genet. 84:97-105

Gill KS, Lubbers EL, Gill BS, Raupp WJ, Cox TS (1991) A genetic-linkage map of Triticum tauschii (DD) and its relationship to the D-genome of bread wheat (AABBDD). Genome 34:362–374

Goff, SA, Ricke D, Lan TH, Presting G, Wang RL, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchinson D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong JP, Miguel T, Paszkowski U, Zhang SP, Colbert M, Sun WL, Chen LL, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu YS, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92-100

Griffiths S, Sharp R, Foote TN, Bertin I, Wanous M, Reader S, Colas I and Moore G (2006) Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat. Nature 439:749-752

Gu Y, Huo N, Vogel J, Garvin D, Lazo G, Anderson O (2006) Construction and characterization of two BAC libraries for *Brachypodium distachyon*, a new model for grass genomics. In: Plant & Animal Genomes XIV Conference, San Diego (CA)

Gupta PK, Balyan HS, Edwards KJ, Isaac P, Korzun V, Röder M, Gautier MF, Joudrier P, Schlatter AR, Dubcovsky J, De la Pena RC, Khairallah M, Penner G, Hayden MJ, Sharp P, Keller B, Wang RCC, Hardouin JP, Jack P, Leroy P (2002) Genetic mapping of 66 new microsatellite (SSR) loci in bread wheat. Theor Appl Genet 105:413–422

Guyomarc'h H, Sourdille P, Edwards KJ, Bernard M (2002) Studies of the transferability of microsatellites derived from *Triticum tauschii* to hexaploid wheat and to diploid related species using amplification, hybridization and sequence comparisons. Theor Appl Genet 105:736-744

Guyot R, Yahiaoui N, Feuillet C, Keller B (2004) In silico comparative analysis reveals a mosaic conservation of genes within a novel colinear region in wheat chromosome 1AS and rice chromosome 5S. Funct Integr Genomics 4:47-58

Guyot R, Keller B (2004) Ancestral genome duplication in rice. Genome 47:610-614

Han F, Kleinhofs A, Ullrich SE, Kilian A, Yano M and Sasaki T (1998) Synteny with rice: analysis of barley malting quality QTLs and *rpg4* chromosome regions. Genome 41:373-380

Hasterok R, Marasek A, Donnison IS, Armstead I, Thomas A, King IP, Wolny E, Idziak D, Draper J, Jenkins G (2006) Alignment of the genomes of *Brachypodium distachyon* and temperate cereals and grasses

using bacterial artificial chromosome landing with fluorescence in situ hybridization. Genetics 173:349-362

Horvath H, Rostoks N, Brueggeman R, Steffenson B, von Wettstein D and Kleinhofs A (2003) Genetically engineered stem rust resistance in barley using the *Rpg1* gene. Proc Natl Acad Sci USA 100:364-369

Huang XQ, Borner A, Roder MS, Ganal MW (2002) Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. Theor Appl Genet 105:699-707

Huang L, Brooks SA, Li WL, Fellers JP, Trick HN, Gill BS (2003) Map-based cloning of leaf rust resistance gene *Lr21* from the large and polyploid genome of bread wheat. Genetics 164:655-664

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793-800.

Isidore E, Scherrer B, Chalhoub B, Feuillet C, Keller B (2005) Ancient haplotypes resulting from extensive molecular rearrangements in the wheat A genome have been maintained in species of three different ploidy levels. Genome Res 15:526-536

Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR (2004) Pack-MULE transposable elements mediate gene evolution in plants. Nature 431:569-573

Joshi AK, Chand R, Kumar S, Singh RP (2004) Leaf tip necrosis: a phenotypic marker associated with resistance to spot blotch disease in wheat. Crop Sci 44:792-796

Kalendar R (2006) FastPCR, PCR primer design, DNA and protein tools, repeats and own database searches program. www.biocenter.helsinki.fi/bi/programs/fastpcr.htm

Kaplinsky NJ, Braun DM, Penterman J, Goff SA, Freeling M (2002) Utility and distribution of conserved noncoding sequences in the grasses. Proc Natl Acad Sci USA 99:6147-6151

Katari MS, Balija V, Wilson RK, Martienssen RA McCombie WR (2005) Comparing low coverage random shotgun sequence data from *Brassica oleracea* and *Oryza sativa* genome sequence for their ability to add to the annotation of *Arabidopsis thaliana*. Genome Res 15:496-504

Kaur M, Saini RG, Preet K (2000). Adult plant leaf rust resistance from 111 wheat (*Triticum aestivum* L.) cultivars. Euphytica 113:235-243

Keller B, Feuillet C, Yahiaoui N (2005) Map-based isolation of disease resistance genes from bread wheat: cloning in a supersize genome. Genet Res 85:93-100

Keller B, Feuillet C (2000). Colinearity and gene density in grass genomes. Trends Plant Sci 5:246-251

Kerber ER, Aung T (1999). Leaf rust resistance gene *Lr34* associated with nonsuppression of stem rust resistance in the wheat cultivar Canthatch. Phytopathology 89:518-521

Kloppers FJ, Pretorius ZA (1997) Effects of combinations amongst genes *Lr13, Lr34* and *Lr37* on components of resistance in wheat to leaf rust. Plant Pathol 46:737-750

Kolmer JA (1996) Genetics of resistance to wheat leaf rust. Annual Review of Phytopathology 34:435-455

Kolmer JA, Liu JQ (2001) Simple inheritance of partial resistance to leaf rust in two wheat cultivars. Plant Pathol 50:546-551

Kurata N, Moore G, Nagamura Y, Foote T, Yano M, Minobe Y, Gale M (1994). Conservation of genome structure between rice and wheat. Bio-Technology 12:276-278

Lander ES, Green P, Abrahamson J, Barlow A, Daley MJ, Lincoln SE, Newburg L (1987) MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181

Laurie DA, Devos KM (2002) Trends in comparative genetics and their potential impacts on wheat and barley research. Plant Mol Biol 48:729-740

Leister D, Kurth J, Laurie DA, Yano M, Sasaki T, Devos K, Graner A, Schulze-Lefert P (1998) Rapid reorganization of resistance genes homologoues in cereal genomes. Proc Natl Acad Sci USA 95:370-375

Li ZK, Luo LJ, Mei HW, Paterson AH, Zhao XZ, Zhong DB, Wang YP, Yu XQ, Zhu L, Tabien R, Stansel JW, Ying CS (1999) A "defeated" rice resistance gene acts as a QTL against a virulent strain of *Xanthomonas oryzae* pv. *oryzae*. Mol Gen Genet. 261:58-63

Liu JQ, Kolmer JA (1998) Genetics of stem rust resistance in wheat cvs. Pasqua and AC Taber. Phytopathology 88:171-176

Luo MC, Thomas C, You FM, Hsiao J, Ouyang S, Buell CR, Malandro M, McGuire PE, Anderson OD, Dvorak J (2003) High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics 82:378-389

Ma H, Singh RP (1996) Contribution of adult plant resistance gene *Yr18* in protecting wheat from yellow rust. Plant Dis 80:66–69

Ma JX and Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes Proc Nat Acad Sci USA 101:12404-12410

Ma J, SanMiguel P, Lai J, Messing J, Bennetzen JL (2005) DNA rearrangement in orthologous *orp* regions of the maize, rice and sorghum genomes. Genetics 170:1209-1220

McFadden ES, Sears ER (1946) The origin of *Triticum spelta* and its free-threshing hexaploid relatives. J Hered 37:81-89, 107-116

McIntosh RA (1992) Close genetic linkage of genes conferring adult-plant resistance to leaf rust and stripe rust in wheat. Plant Pathol 41:523–527

McIntosh RA, Wellings CR and Park RF (1995) Wheat rusts: an atlas of resistance genes. CSIRO, Australia, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Messmer MM, Keller M, Zanetti S, Keller B (1999) Genetic linkage map of a wheat x spelt cross. Theor Appl Genet 98:1163-1170

Messmer MM, Seyfarth R, Keller M, Schachermayr G, Winzeler M, Zanetti S, Feuillet C, Keller B (2000) Genetic analysis of durable leaf rust resistance in winter wheat. Theor Appl Genet 100:419-431

Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A (2005) Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nat Genet 37:997-1002

Nelson JC, Singh RP, Autrique JE, Sorrells ME (1997) Mapping genes conferring and suppressing leaf rust resistance in wheat. Crop Sci 37:1928-1935

Nelson JC, Sorrells ME, VanDeynze AE, Lu YH, Atkinson M, Bernard M, Leroy P, Faris JD, Anderson JA. (1995) Molecular mapping of wheat - Major genes and rearrangements in homoeologous group 4, group 5, and group 7. Genetics 141:721-731

Niks RE, Rubiales D (2002) Potentially durable resistance mechanisms in plants to specialised fungal pathogens. Euphytica 124:201-216

Nilmalgoda SD, Cloutier S, Walichnowski AZ (2003) Construction and characterization of a bacterial artificial chromosome (BAC) library of hexaploid wheat (*Triticum aestivum* L.) and validation of genome coverage using locus-specific primers. Genome 46:870-878

Oerke E-C, Dehne H-W (2004) Safeguarding production-losses in major crops and the role of crop protection. Crop Prot 23:275-285

Ogbonnaya FC, Halloran GM, Lagudah ES (2005) D genome of wheat – 60 years from Kihara, Sears and McFadden. In: Tsunewaki K (ed) Frontiers of wheat bioscience. Kihara memorial foundation for the advancement of life sciences, Yokohama, pp 205-220

Paillard S, Schnurbusch T, Tiwari R, Messmer M, Winzeler M, Keller B, Schachermayr G (2004) QTL analysis of resistance to *Fusarium* head blight in Swiss winter wheat (*Triticum aestivum* L.) Theor Appl Gen 109:323-332

Paillard S, Schnurbusch T, Winzeler M, Messmer M, Sourdille P, Abderhalden O, Keller B, Schachermayr G (2003) An integrative genetic linkage map of winter wheat (*Triticum aestivum* L.). Theor Appl Genet 107:1235-1242

Parlevliet JE (1979) Components of resistance that reduce the rate of epidemic development. Ann Rev Phytopathol 17:202-222

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals and its consequences for comparative genomics. Proc Natl Acad Sci USA, 101:9903-9908

Paux E, Sourdille P, Salse J, Leroy P, Bartos J, Mestiri I, Chalhoub B, Quetier F , Bernard M, Dolezel J, Feuillet C (2006) Chromosome based strategies to decipher the hexaploid wheat genome: chromosome 3B, a case study. In: 14th international plant and animal genome conference, San Diego,USA. Workshop 46

Pestsova E, Ganal MW, Roder MS (2000) Isolation and mapping of microsatellite markers specific for the D genome of bread wheat. Genome 43:689-697

Prasad M, Varshney RK, Roy JK, Balyan HS, Gupta PK (2000) The use of microsatellites for detecting DNA polymorphism, genotype identification and genetic diversity in wheat. Theor Appl Genet 100:584-592

Qi LL, Echalier B, Friebe B, Gill BS (2003) Molecular characterization of a set of wheat deletion stocks for use in chromosome bin mapping of ESTs. Funct Integr Genomics 3:39-55

Rice P, Longden I, Bleasby A (2000) EMBOSS: The European molecular biology open software suite. Trends Genet 16:276-277

Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier M-H, Leroy P, Ganal MW (1998) A microsatellite map of wheat. Genetics 149:2007-2023

Rubiales D, Niks RE (1995) Characterization of *Lr34*, a major gene conferring nonhypersensitive resistance to wheat leaf rust. Plant Dis 79:1208-1212

Sakata K, Nagamura Y, Numa H, Antonio BA, Nagasaki H, Idonuma A, Watanabe W, Shimizu Y, Horiuchi I, Matsumoto T, Sasaki T, Higo K (2002) RiceGAAS: an automated annotation system and database for rice genome sequence. Nucleic Acids Res 30:98-102

Sambrook J, Russel DW (2001) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratoy Press, Cold Spring Harbor, NY, USA

SanMiguel PJ, Ramakrishna W, Bennetzen JL, Busso C, Dubcovsky J (2002) Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome 5Am. Funct Integr Genomics 2:70-80

SAS Institute 1991. SAS© language and procedures: Usage 2, Version 6, First Edition. SAS Institute Inc., Cary, NC, USA

Sasaki T, Matsumoto T, Yamamoto K, Sakata K, Baba T, Katayose Y, Wu JZ, Niimura Y, Cheng ZK, Nagamura Y, Antonio BA, Kanamori H, Hosokawa S, Masukawa M, Arikawa K, Chiden Y, Hayashi M, Okamoto M, Ando T, Aoki H, Arita K, Hamada M, Harada C, Hijishita S, Honda M, Ichikawa Y, Idonuma A, Iijima M, Ikeno M, Ito S, Ito T, Ito Y, Iwabuchi A, Kamiya K, Karasawa W, Katagiri S, Kikuta A, Kobayashi N, Kono I, Machita K, Maehara T, Mizuno H, Mizubayashi T, Mukai Y, Nagasaki H, Nakashima M, Nakama Y, Nakamichi Y, Nakamura M, Namiki N, Negishi M, Ohta I, Ono N, Saji S, Sakai K, Shibata M, Shimokawa T, Shomura A, Song JY, Takazaki Y, Terasawa K, Tsuji K, Waki K, Yamagata H, Yamane H, Yoshiki S, Yoshihara R, Yukawa K, Zhong HS, Iwama H, Endo T, Ito H, Hahn JH, Kim, H. I., Eun, M. Y., Yano, M., Jiang, J. M. and Gojohori, T. 2002. The genome sequence and structure of rice chromosome 1. Nature 420:312-316

Sawhney RN, Sharma JB (1999) Novel complementary genes for adult plant leaf rust resistance in a wheat stock carrying the 1BL-1RS translocation. Plant Breeding 118:269-271

Sawhney RN (1992) The role of *Lr34* in imparting durable resistance to wheat leaf rust through gene interaction. Euphytica 61:9-12

Schnurbusch T, Bossolini E, Messmer M, Keller B (2004b) Tagging and validation of a major quantitative trait locus for leaf rust resistance and leaf tip necrosis in winter wheat cultivar Forno. Phytopathology 94:1036-1041

Schnurbusch T, Paillard S, Schori A, Messmer M, Schachermayr G, Winzeler M, Keller B (2004a) Dissection of quantitative and durable leaf rust resistance in Swiss winter wheat reveals a major resistance QTL in the *Lr34* chromosomal region. Theor Appl Genet 108:477-484

Sears ER (1954) The aneuploids of common wheat. Mo Agr Exp Sta Res Bull 572:1–59

Sears ER (1966a) Nullisomic-tetrasomic combinations in hexaploid wheat, pp. 29–45 in Chromosome Manipulation and Plant Genetics, edited by R. RILEY and K. R. LEWIS. Oliver & Boyd, Edinburgh.

Sears ER (1966b) Chromosome mapping with the aid of telocentrics. In: MacKey J (ed) 2nd Wheat Genet Symp, Lund, pp 370–381

Shen B, Wang DM, McIntyre CL, Liu CJ (2005) A "Chinese Spring" wheat (*Triticum aestivum* L.) bacterial artificial chromosome library and its use in the isolation of SSR markers for targeted genome regions. Theor Appl Genet 111:1489-1494

Singh RP (1992a) Association between gene *Lr34* for leaf rust resistance and leaf tip necrosis in wheat. Crop Sci 32:874-878

Singh RP, Huerta-Espino J (1997) Effect of leaf rust resistance gene *Lr34* on grain yield and agronomic traits of spring wheat. Crop-Science 37:390-395

Singh RP, Huerta-Espino J, William HM (2005) Genetics and breeding for durable resistance to leaf and stripe rusts in wheat. Turk J Agric For 29:121-127

Singh RP, Huerta-Espino J. (2003) Effect of leaf rust resistance gene *Lr34* on components of slow rusting at seven growth stages in wheat. Euphytica 129:371-376

Singh RP, Kumar N, Bandopadhyay R, Rustgi S, Sharma S, Balyan HS, Gupta PK (2006) Development and use of anchored-SSRs to study DNA polymorphism in bread wheat (*Triticum aestivum* L.). Mol Ecol Notes 6:296-299

Singh RP (1992b) Genetic association of leaf rust resistance gene *Lr34* with adult plant resistance to stripe rust in bread wheat. Phytopathology 82:835-838

Singh RP (1993) Genetic association of gene *Bdv1* for tolerance to barley yellow dwarf virus with genes *Lr34* and *Yr18* for adult plant resistance to rusts in bread wheat. Plant Dis 77:1103-1106

Singh RP (1993) Resistance to leaf rust in 26 Mexican wheat cultivars. Crop Sci 33:633-637

Singh RP, Gupta AK (1991) Genes for leaf rust resistance in Indian and Pakistani wheats tested with Mexican pathotypes of *Puccinia recondita* f.sp. *tritici*. Euphytica 57:27-36

Singh RP, Rajaram S (1992) Genetics of adult plant resistance of leaf rust in Frontana and 3 CIMMYT wheats. Genome 35:24-31

Singh RP, Mujeeb-Kazi A, Huerta-Espino J (1998) *Lr46*: A gene conferring slow-rusting resistance to leaf rust in wheat. Phytopathology 88:890-894

Somers DJ, Isaac P, Edwards K (2004) A high density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). Theor Appl Genet 109:1105–1114

Song Q, Shi J, Singh S, Fickus E, Costa J, Lewis J, Gill B, Ward R, Cregan P (2005) Development and mapping of microsatellite (SSR) markers in wheat. Theor Appl Genet 110:550-560

Song R, Llaca V, Messing J (2002) Mosaic organization of orthologous sequences in grass genomes. Genome Res 12:1549-1555

Sonnhammer EL, Durbin R (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene 167:1-10

Sorrells ME, La Rota M, Bermudez-Kandianis CE, Greene RA, Kantety R, Munkvold JD, Miftahudin T, Mahmoud A, Ma X, Gustafson PJ, Qi LL, Echalier B, Gill BS, Matthews DE, Lazo GR, Chao S, Anderson OD, Edwards H, Linkiewicz AM, Dubcovsky J, Akhunov ED, Dvorak J, Zhang D, Nguyen

HT, Peng J, Lapitan NLV, Gonzalez-Hernandez JL, Anderson JA, Hossain K, Kalavacharla V, Kianian SF, Choi D-W, Close TJ, Dilbirligi M, Gill KS, Steber C, Walker-Simmons MK, McGuire PE, Qualset CO (2003) Comparative DNA sequence analysis of wheat and rice genomes. Genome Res 13:1818-1827

Spielmeyer W, McIntosh RA, Kolmer J, Lagudah ES (2005) Powdery mildew resistance and *Lr34/Yr18* genes for durable resistance to leaf and stripe rust cosegregate at a locus on the short arm of chromosome 7D of wheat. Theor Appl Genet 111:731-735

Stein N, Feuillet C, Wicker T, Schlagenhauf E, Keller B (2000) Subgenome chromosome walking in wheat: A 450-kb physical contig in *Triticum monococcum* L. spans the *Lr10* resistance locus in hexaploid wheat (*Triticum aestivum* L.). Proc Natl Acad Sci USA 97:13436-13441

Suenaga K, Singh RP, Huerta-Espino J, William HM (2003) Microsatellite markers for genes *Lr34/Yr18* and other quantitative trait loci for leaf rust and stripe rust resistance in bread wheat. Phytopathology 93:881-890

Swigonova Z, Bennetzen JL, Messing J (2005) Structure and evolution of the r/b chromosomal regions in rice, maize and sorghum. Genetics 169:891-906

Tarchini R, Biddle P, Wineland R, Tingey S, Rafalski A (2000) The complete sequence of 340 kb of DNA around the rice *Adh1-adh2* region reveals interrupted colinearity with maize chromosome 4. Plant Cell 12:381-391

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W, improving the sensitivity of progressive multiple sequence alignment through sequence weighting., position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673-4680.

Utz HF, Melchinger AE (2000) PLABQTL: A computer program to map QTL. Version 1.1, Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, Stuttgart, Germany

van der Plank JE (1963) Plant Diseases: epidemics and control. Academic Press, New York, USA

VanDeynze AE, Nelson JC, Yglesias ES, Harrington SE, Braga DP, McCouch SR, Sorrells ME (1995) Comparative mapping in grasses - wheat relationships. Mol Gen Genet 248:744-754

Vogel JP, Raab TK, Schiff C, Somerville SC (2002) *PMR6*, a pectate lyase-like gene required for powdery mildew susceptibility in *Arabidopsis*. Plant Cell 14:2095-2106

Vogel, J.P., Gu, Y.Q., Twigg, P., Lazo, G.R., Laudencia-Chingcuanco, D., Hayden, D.M., Donze, T.J., Vivian, L.A., Stamova, B. and Coleman-Derr, D. (2006) EST sequencing and phylogenetic analysis of the model grass *Brachypodium distachyon*. Theor Appl Genet 113:186-195

Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. J Hered 93:77-78

White SE and Doebley JF (1999) The molecular evolution of terminal ear 1, a regulatory gene in the genus *Zea*. Genetics 153:1455-1462

Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu ZD, Dubcovsky J, Keller B (2003) Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and A(m) genomes of wheat. Plant Cell 15:1186-1197

Wicker T, Guyot R, Yahiaoui N, Keller B (2003a) CACTA transposons in Triticeae. A diverse family of high-copy repetitive elements. Plant Physiol 132:52-63

Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E, Keller B (2001). Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J 26:307-316

Winzeler M, Mesterhazy A, Park RF, Bartos P, Csosz M, Goyeau H, Ittu M, Jones E, Loschenberger F, Manninger K, Pasquini M, Richter K, Rubiales D, Schachermayr G, Strzembicka A, Trottet M, Unger O, Vida G, Walther U (2000) Resistance of European winter wheat germplasm to leaf rust. Agronomie 20:783-792

Wolfe KH, Gouy M, Yang Y-W, Sharp PM, Li W-H (1989) Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc Natl Acad Sci USA 86:6201-6205

9. ACKNOWLEDGMENTS

I am grateful to…

Prof. Dr. Beat Keller for his continous support and confidence throughout my thesis.

Prof. Dr. Enrico Martinoia and Prof. Dr. Ueli Grossniklaus for taking charge of the co-referate.

Simon Krattinger, Dr. Thorsten Schnurbusch, Dr. Thomas Wicker, Philip Knobel and Bea Senger for their outstanding engagement throughout my thesis. Large part of this work would not have been realized without their continuous help.

All present and past members of Beat Keller's group, my family and all my friends in Zürich. Your direct and indirect support have made this possible.

10. CURRICULUM VITAE

Surname	BOSSOLINI
First name	Eligio
Date and place of birth	17th April 1978, Bozzolo (MN) Italy
Nationality	Italian
Education	
Abitur 1997	Diploma di maturita' scientifica (44/60), Liceo Scientifico 'P.A. Maggi' in Viadana (MN) Italy
University studies 1997-2002	University diploma (Laurea) in Plant Biotechnology at the University 'Alma mater studiorum' of Bologna (Italy) in 2002 with a score of 110+/110. The University diploma thesis is entitled 'QTL analysis of morphophysiological traits in a interspecific cross-derived progeny'.
Other colleges 2001-2002	One academic year at the University of Minnesota, United States, as exchange student in the department of Agronomy and Plant Genetics for the University diploma thesis. Topic was mapping and QTL analysis in barley interspecific backcross. The project was supervised by Prof. Dr. G. Muehlbauer and Prof. Dr. Roberto Tuberosa.
Doctorate studies 2003-2006	PhD in plant molecular biology; Title of PhD thesis: 'Marker development for the durable leaf rust resistance gene *Lr34* of wheat using sequence information from rice, *Aegilops tauschii* and *Brachypodium sylvaticum*' supervised by Prof. Beat Keller, University of Zürich. During my PhD I had the opportunity to supervise two master students and I reviewed several papers.
Working experience besides the PhD employment	Seasonal jobs in Italy (1995-2001). Employed by the Department of Agronomy and Plant Genetics at the University of Minnesota, US (2001-2002).